DOCUMENT RESUME

ED 168 815 SE 026 588

AUTHOR	Blake, Anthony J. D., Ed.; Power, Colin N., Ed.
TITLE	Research in Science Education, Volume 7. Proceedings of the Annual Conference of the Australian Science Education Research Association (8th, Riverina College of Advanced Education, Wagga Wagga, New South Wales, May 16-18, 1977).
INSTITUTION	Australian Science Education Research Association.
PUB DATE	77
NOTE	233p.; For volume 6, see ED 143 501 ; May be marginally legible due to some small print
EDRS PRICE	MF01/PC10 Plus Postage.
DESCRIPTORS	College Science; *Conference Reports; *Educational Research; Elementary School Science; Elementary Secondary Education; Higher Education; *Instruction; *Science Education; Science Teachers; *Secondary School Science
IDENTIFIERS	*Research Reports

ABSTRACT

This volume contains papers presented at the eighth Annual Conference of the Australian Science Education Research Association held at Riverina College of Advanced Education in May, 1977. Paper topics include: (1) the nature of science problems; (2) logical connectives in science; (3) a model of cognitive processes; (4) measures of meaningful learning of intellectual skills; (5) primary school pupils' ability to see scientific problems in everyday phenomena; (6) an evaluation of the implementation of the junior secondary science syllabus in N.S.W.; (7) school factors affecting curriculum change in science; (8) the systems approach and its application to curriculum research; (9) formal schooling and the attitude of Papua New Guineans towards natural phenomena; (10) suggestions from China on the relevance of science instruction; (11) a student's eye view of the biology classroom; (12) teacher characteristics and pupil outcomes in secondary science classrooms; and (13) sex differences in student perceptions for and perceptions of learning outcomes and classroom activities in year-eleven biology. (BB)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED168815

U S DEPARTMENT OF HEALTH
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Colin N. Power

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) AND USERS OF THE ERIC SYSTEM

SE 026 588

2

RESEARCH IN SCIENCE EDUCATION

VOLUME 7

Proceedings of the Eighth Annual Conference of the Australian Science Education Research Association, Riverina College of Advanced Education, Wagga Wagga, N.S.W., May 16-18, 1977.

Edited by: *Anthony J. D. Blake, Riverina College of Advanced Education*

General Editor: *Colin N. Power, University of Queensland*

Editorial Board: Arthur Lucas, Flinders University
Max Maddock, University of Newcastle
Leo West, Monash University
Richard White, Monash University

Business Manager: Jim Butler, University of Queensland

AUSTRALIAN SCIENCE EDUCATION RESEARCH ASSOCIATION

ERRATUM
page 9 and page 131 have been interchanged in the printing process

All correspondence concerning this publication should be addressed to Prof. Colin N. Power, School of Education, Flinders University, Bedford Park, 5042, Australia.

Copyright © 1977 by Australian Science Education Research Association.
Published by Australian Science Education Research Association.
Printed by the University of Queensland Printery.

CONTENTS

Page

Preface (v)

B. C. Newman
The nature of science problem 7

P. L. Gardner
Logical connectives in science: A summary of the findings 9

Richard T. White
A model of cognitive processes 25

N. C. Kellett, L. H. T. West and C. Woodruff
Measures of meaningful learning of intellectual skills 33

David J. Symington
Primary school pupils' ability to see scientific problems in everyday phenomena 41

Sue McBurney
An evaluation of the implementation of the junior secondary science syllabus in N.S.W. 51

John M. Owen
School factors affecting curriculum change in science 61

Roger J. Osborne and Peter S. Freyberg
The systems approach and its application to curriculum research 71

M. N. Maddock
Formal schooling and the attitude of Papua New Guineans towards natural phenomena – a replication and extension study 83

Frederick J. Thomas
Suggestions from China on the "relevance" of science instruction 97

Effie D. Best
A student's eye view of the biology classroom: Teacher, activities and outcomes. 105

Jeffrey R. Northfield and Barry J. Fraser
Teacher Characteristics and pupil outcomes in secondary science classrooms. 113

Janet V. Keightley
Sex differences in student perceptions for, and perceptions of, learning outcomes and classroom activities in Year 11 biology 123

Contents - contd.

Page

M. J. Sullivan
Critical thinking in science – some problems in measurement 131

P. P. Lynch
Laboratory-based teaching in physical science and its relation to pupil attainment. 139

Renato Schibeci
Attitudes to science: A semantic differential instrument. 149

Roger J. Osborne
Subjectively assessed student profiles in physics 157

J. K. Gilbert
The study of student misunderstandings in the physical sciences. 165

P. P. Lynch and G. C. Gerrans
The aims of first year chemistry courses, the expectation of new students and subsequent course influences. 173

C. B. Kings
Social characteristics, motivations, achievements and choices of H.S.C. subjects of environmental studies in a tertiary institution – some preliminary findings. 181

R. Lewis
Transition from Year 12 schooling to the workfoce: Does scientific achievement help? 193

R. A. Fawns
Student teacher questioning in two enquiry directed settings of science teaching. 203

John Ainley
The impact of science facilities on science learning. 213

Geoffrey W. Beeson
The application of theoretical research findings to a practical learning problem. 223

Lindsay D. Mackay and Richard T. White
Changes in pupils' perceptions of scientists, science teachers and self. 231

Contributors 239

PREFACE

The papers published in this journal represent the majority of papers presented at the 8th Annual Conference of the Australian Science Education Research Association. It was the second occasion that the conference was held at a provincial venue and the first time that the host institution was a College of Advanced Education.

As twenty-eight formal research papers were presented at the conference, authors were requested to constrain the length of papers so that the volume stayed within reasonable size limits. However, every effort has been made to ensure that each paper published is in accord with the conference presentation.

The relatively large number of contributions at the conference ranged over a wide area of interests in science education. White introduced a thought-provoking model of cognitive processes which has implications well beyond science education. Osborne and Freyberg discussed a systems approach to curriculum research which similarly has broad implications. On the other hand, many other contributors concentrated on specific topics of concern in science education. Amongst these papers one will note some concentration upon issues associated with science education at secondary and tertiary levels. Conversely, only one paper dealt with research into aspects of primary science education. It is to be hoped that the increased interest in science in the primary curriculum will be supported by a greater degree of research activity in the area in the future.

Anthony J. D. Blake
School of Teacher Education
Riverina College of
Advanced Education.

Research in Science Education, 1977, 7, 1-7.

THE NATURE OF SCIENCE PROBLEM

B. C. Newman

Introduction

The well known Test on Understanding Science (W) developed by Cooley and Klopfer (1961) is based upon an understanding of the nature of science and scientists which recognises a number of themes grouped into three major areas: The scientific enterprise, scientists, and the methods and aims of science. Investigations using this measure often report analyses based upon total scores and subscores for these areas. Now one may object to this way of dividing up the scientific cake and the question arises: What conceptions of the thing called science are allowable and according to what categories can it legitimately be analysed? In this connection, with reference to the Test on Understanding Science (W), one may ask the following questions. Why is 'the nature of science' not inclusive of 'the nature of scientists'? Why should 'the scientific enterprise' made up of themes such as 'scientific societies' and 'interaction of science and society' be considered somewhat distinct from 'scientists' made up of themes such as 'institutional pressures on scientists'? Why should both areas be considered to be somewhat distinct from 'the methods and aims of science' made up of themes such as 'tactics and strategy of sciencing' and 'science and technology'?

The Existence of Differences in the Conception and Categorisation of 'the Nature of Science'

A glance at the bare descriptions given to just a few of the instruments designed for the evaluation of people's understanding of 'the nature of science' or some aspect of it, suggests that the variation in the conception and categorisation of 'the nature of science' is considerable. The Wisconsin Inventory of Science Processes (Pella, 1967) is concerned with the "assumptions, activities, objectives and products of science". The three subscales of the Test on the Social Aspects of Science (Korth, 1968) deal with the interaction that exists among science, technology and society, the social nature of the scientific enterprise and the social and political responsibilities of scientists. The Welch Science Process Inventory (D) (Welch, 1969) deals with "the activities, assumptions, products and ethics of science". Aikenhead (1973) in comparing six such instruments evaluated them in a qualitative manner, from the point of view of the extent to which they dealt with: knowledge of subject matter, content and skills in carrying out scientific processes, understanding about the processes of science, understanding the nature of science, understanding the relationships between science and society, and attitudes towards science (perhaps some of these can be ignored, being considered as not being concerned with the nature of science itself). A large variation obviously exists in descriptions of science and presumably in the conceptions and categorisations of science to which people feel they can legitimately subscribe. Such variation might well be expected to increase even further when it comes to matters of detail.

A Correct Understanding of 'the Nature of Science'?

Now it would be unhelpful to suggest that nonetheless, in spite of this large variation, what we are looking for and ultimately hope to find is the correct understanding of 'the nature of science' and the one and only legitimate way of characterising it. Such a quest assumes that the time would arrive when serious debate on such matters would cease. The history of the philosophy of science holds out no such hope.

Nor, as Lucas (1975) points out, is an adequate model of science guaranteed by appealing to some panel of so-called experts. What the majority of experts agrees to is not to be confused with what would be regarded as adequate. While any minority, or any majority of a different panel, held out for an alternative view, we would be suspicious that the model entertained by the majority of one panel was inadequate. Who are the experts anyway? It is not just a matter of deciding whether to ask the scientists, or the science educators, or the historians or philosophers of science, or some combination of them, but also of deciding which philosophers of science, which scientists, etc.

A look at what happened in an attempt to provide partial validation for the Science Understanding Measure (Coxhead and Whitfield, 1975) is instructive. Of the 31 experts to whom a questionnaire was sent, only 19 gave replies which either made no adverse comments or made comments that were taken into account in the revision of the measure. Of the remainder, some did not reply, or did but did not answer the questionnaire, or had or appeared to have fundamental objections to the measure itself. It is quite problematical what the results would have been if a different sample of experts had have been chosen and it is understandable if one has little confidence in the validity of a measure grounded in such procedures.

Various Possible Understandings of 'the Nature of Science'

Simply put, any statement outlining 'the nature of science' will always have a question mark over it. This is not to say, however, that one should either give up making such statements or not argue for them when they are made. Science Educators cannot be neutralists in this matter. They will have some view of the nature of science and here and there it will permeate what they do in science education. Nor, we would hope, would we be happy to abandon attempts to justify any position we hold on this issue. Indeed, in order to better justify a particular view, it would seem beneficial to be aware of a variety of views that one could adopt.

In order to illustrate some of the various possibilities open to us, I will proceed by asking a number of questions of the nature of science to which I will give rather simple answers, followed by brief comments on these answers with indications of other possible answers. Depending upon what one's initial 'feel' for the nature of science is, what one's diffuse conception of it is, only some questions can be considered legitimate and answerable and some questions can be considered as more important than others. I have chosen a broad conception to permit a number of questions of a general nature to be asked. It cannot be guaranteed, however, that some will not argue that some of the questions to follow cannot be properly asked of the nature of science and/or that other questions, not asked, should have been. Even granted, however, that the questions are legitimate and important, the variation in views that could still exist would be reflected in the answers and our attention is centred on some such possible answers and the questions from which they arise.

Q. What are the basic ingredients of science and how are they related?

A. Labour, Content, Processes and Products. Scientific labour is expended on scientific content using scientific processes to produce scientific products.

This type of answer suggests that appeal is being made to a production-type model, where labour works with raw materials using certain means of production for their transformation into products, a model congruent with the elaboration by Althusser (1969, p.166) of the notion of 'practice'. Now Althusser views the raw material of what he terms 'scientific theoretical practice' as consisting of "ideological concepts . . . scientific 'facts' . . . (and) scientifically elaborated concepts" and the product of such practice to consist of "the corpus of concepts, whose more or less contradictory unity constitutes the 'theory' of the science at the time" (p.184). Such a model is suggestive of how various disciplines of science might be distinguished. Their raw materials and products particularly, but perhaps also their means of production could be appealed to in distinguishing chemistry from physics, etc. It is not suggestive in itself of a reductionist relationship among such disciplines.

Now it is of significance that facts are placed in inverted commas, because, for Althusser and Balibar (1970), there are no such things as facts in the realm of knowledge, rather products of thought which are to be distinguished from the real objects in the real world to which they relate (p.41).

Furthermore, given an Althusserian interpretation of the reply, the perhaps expected reply of 'scientific concepts' and 'scientific methods', by comparison would not cause us to focus as much on the changing nature of the content of science and certainly would direct less attention to those who apply the methods. While scientists do receive attention in the answer given, society and any role it might play in, for example, determining the extent and nature of the labour, content, processes and products, does not. As it stands, it could be argued, the reply is quite deficient in this respect.

Q. What does science achieve and how does it achieve it?

A. By developing more embracive theories with greater confirmation it enables better explanations and predictions to be made and better control to be exercised.

Neither the question nor the answer necessarily presuppose that scientists have to aim at what they achieve, but it suggests it, and so both the question and the answer could be considered to be dealing with scientific aims. Some may argue for the exclusion of 'control' as an aim, claiming that better control is automatically possible with better explanations and predictions and, furthermore, control is aimed at by those applying science rather than by those doing it. Scheffler (1964) would see explanation and prediction as too limiting, since he recognises two significant types of aims of science that he terms 'substanting' and 'positing' which, though including respectively 'explaining' and 'predicting', refer to other sets of activities as well (pp.46-57). It is presumably of significance that the answer to the question does not include a reference to such as 'searching for the truth' or 'discovering the laws of nature'.

As regards how these aims are achieved, the answer focuses attention upon theories, the confirmation and hence the testing of theories, theory development, theory comparison and theory replacement. If 'confirmed' were replaced by 'not refuted', we would have a position very similar to that of Popper (1972 (a), pp.13-16). However, it is Lakatos (1970) who gives such great detailed attention to theories, their development and their replacement, and for him, confirmation is of considerable importance. On the other hand, Feyerabend (1970), who gives prominence to incommensurable theories and downgrades the importance of methodological rules, contrary to Popper and Lakatos, adopts a fundamentally different stance. Propagate theories no matter what their relationship with other theories and the evidence and don't readily relinquish them, is the hallmark of his position. Kuhn (1970), who likewise sees incommensurable theories as significant, adopts yet another position. It is the paradigm of disciplinary matrix, a more complex unit than the simple theory, though inclusive of theories, together with the scientific revolution in which a paradigm changes, which are seen as central to the development of scientific thought. It is perhaps in this area, that of scientific methodology, that debate among philosophers of science is most striking.

Q. Who does science and how do they do it?

A. Scientific communities with their scientific attitudes and aims impose standards on the individuals of those communities.

The question and answer are not to be seen as clearly divorced from the previous question and answer (or even the earlier set), although in this case greater attention is given to the scientists who practise science, particularly to the notion of 'scientific communities'. Now for Kuhn, the idea of scientific communities is a very important one. A scientific community, consisting of the group of scientists working in a particular area, is largely responsible for the occurrence of a scientific revolution where one paradigm is exchanged for another. It is the scientific community which finally determines for the individual scientist the nature of a scientific problem and how it can be solved. The answer given above, however, may be taken to suggest that in addition, scientific communities, as communities, have attitudes such as scepticism which impose on individual scientists certain patterns of behaviour which take this into account. For example, it could be argued that part of the character of the scientific paper is determined with the scepticism of the scientific community in mind. Furthermore, the answer above suggests that, not only such attitudes but also aims of the scientific community, affect the behaviour of the individuals of that community. If a particular scientific community finds as unacceptable the aim of control over the environment, then the individuals of that community can have no such ostensible aim. It is understandable then how, both in the area of attitudes and of aims, the scientific paper may not be a genuine reflection of the nature of the individual scientist (Medawar, 1963). The distinction between 'private science' and 'public science' held by Holton and Roller (1958, pp.231 232) may be a useful one in this connection.

In opposition to the underlying viewpoint expressed in the answer, however, one may argue that the relationship between a scientific community and the individual scientists of that community is such an exceedingly complex one that a strong emphasis on the effect of the community on the individuals is likely to be grossly misleading. Furthermore, one may even object to the holistic notion of a scientific community altogether and prefer to recognise only individuals having different degrees of influence upon each other.

Q. What are its presuppositions?

A. They are the presuppositions of the individual scientists operating as scientists, at the time considered.

This answer suggests that, as with the objection immediately above, holistic notions are not acceptable, at least in this case, in the area of presuppositions. The inference is that a concise answer cannot be given to the question, that the presuppositions held could be contradictory and numerous, and furthermore, that they may well change with time.

In opposition to this, Popper (1972 (b), pp.252, 253) sees a presupposition concerning the uniformity of nature as a methodological rule to be followed by all scientists, presumably throughout all time. Lakatos (1970) would probably expect such a proposition to form part of the hard core of most theories and consequently the associated research programme, but his position could readily allow many propositions within their hard cores to be the peculiar property of individual scientists. Kuhn (1970, pp.181-191) on the other hand, taking a different position again, sees at least some beliefs and values as belonging to the disciplinary matrix of a scientific community. These presuppositions are those of the community as a community, but as a paradigm changes, so may the presuppositions.

Q. When is science being practised?

A. When falsifiable ideas are proposed and when attempts are made to falsify or confirm them.

This question strikes at the problem of demarcation. When is something science and when is something not science? Provided by 'confirmation' one meant 'corroboration' – a measure of the extent to which the idea, the theory had been tested – the answer given corresponds very well to the falsificationism of Popper (1972 (b)). However for Popper, a theory itself, whether attempts are made to corroborate it or not, is scientific, simply if it is falsifiable, and a theory is falsifiable if a deduction can be made from it, the denial of which is possible by a basic statement derived from an observation. It is of importance to note that for Popper, inductivism is of no consequence in solving the demarcation problem, and in association with this stance he believes that the question of how theories come to be proposed is of no philosophical, only psychological importance.

Lakatos (1970), in opposition to Popper, believes that a theory is only falsified when at least one other theory is available that has excess empirical content over the former theory, some of which excess content has been confirmed. Consequently, it is not really a theory which is appraised as scientific, but a series of theories.

When it comes to Kuhn or Feyerabend, it is not all that clear what makes the scientific endeavour scientific. Feyerabend (1970) does suggest that Kuhn is unable to handle the demarcation problem because he does not tackle the notion of aims and thereby infers that an activity is known to be scientific only by the aim it pursues, perhaps an aim such as seeking explanations.

Now if an activity is scientific if and only if the theory or series of theories associated with it is scientific, *or*, if and only if the aim associated with it is a scientific one, then perhaps one can legitimately describe laws, models or definitions etc. as scientific ones, if and only if they have a strong association with either a scientific theory or series of theories, or a scientific aim. It would follow also that the primary school teacher could perhaps legitimately claim that his pupils were doing science if they were seeking explanations or if they were proposing ideas capable of falsification, regardless of how abstract or complex their facts, laws, models or definitions might be. Conversely, however, it could be argued that some production of so-called scientific apparatus, and the gathering of some so-called scientific data were not scientific because of the lack of significant association with scientific aims or theories. Because of the values attached to the term 'science', it is of concern to discover what can legitimately be labelled as 'science'.

Q What are its relationships with other entities?

A. The nature of scientific practice is in part determined by the economic, political and ideological practice of the society in which it is being practised.

The Althusserian view of the existence of a number of practices in a society and their inter-relationships, gives some support to such an answer. For Althusser (1969) " 'Social practice', the complex unity of the practices existing in a . . . society, contains a large number of distinctive practices" (p.167). These include material (economic), political, ideological and theoretical (including scientific theoretical) practices which transform nature, social relations, men's consciousness, and certain representations, concepts and facts, respectively. Although scientific practice is in one sense autonomous – its aim and the criteria of the adequacy of scientific practice are internal to scientific practice – its nature depends on its "organic relations with other practices which provide (the) sciences with a large proportion of their raw material, and occasionally go as far as to induce more or less profound re-organizations in their theoretical structure" (Althusser and Balibar, 1970, p.60).

All this suggests that science may be different in different societies, although perhaps not in a fundamental way and that in order to understand the science of a society one needs to understand that society, that is, its various practices and their inter-relationships. Consequently, if one understands these practices and sub-practices (e.g. those of technology, education and religion) and recognises how they change, then one is more able to understand how a particular scientific practice came into being, how it has changed and why it has changed; if it has so at all.

Of course, it is open for any one to argue that the science of a society is less or more dependent on the other features of that society than is here indicated. One may argue that science is independent of the society in which it is practised or that science, as we think we know it, because of social effects, is soon to disappear.

Our problem is that science has not yet disappeared and in science education, from time to time, one is inevitably concerned with its characteristics. One of the aims of this paper has been to show that it's of little use to wait around for a definitive answer from, for instance, the philosophers of science, as to what it's all about. There are just too many stances possible, too many questions that could be asked, too much debate about what should be asked and too much variety to what could be answers.

One obvious implication is that those concerned with people learning about the nature of science would be better off being much more precise about what it is they want them to learn and much more precise about how they should learn it. Furthermore, those so concerned would consequently be better off evaluating what people have learned by using measuring instruments specifically designed to cope with the explicit model they are working with. Additionally, as has been maintained, they may be better able to justify their model of the nature of science, if they are more aware of the different ways of analysing science, its many possible facets, the complexities of these facets and the contentious matters that arise in the study of these facets. On the other hand, once Pandora's box has been opened, they might not be able to handle it all and give up the game and throw learning in this area to the wind, or prefer to measure, for instance, the extent to which a learner's model of the nature of science corresponds with someone else's view, without worrying about the justification of either view. But is the responsible educationist happy with such a position?

References

AIKENHEAD, G.S. The measurement of high school students' knowledge about science and scientists. *Science Education*, 1973, 57 (*4*), 539-549.

ALTHUSSER, L. *For Marx*, translated by B. Brewster. Harmondsworth: Penguin, 1969.

ALTHUSSER, L. & BALIBAR, E. *Reading Capital*, translated by B. Brewster. London: NLB, 1970.

COOLEY, W.W. & KLOPFER, L.E. *Test on understanding science: Form W, Manual.* Princeton: Educational Testing Service, 1961.

COXHEAD, P. & WHITFIELD, R.C. *Science understanding measure, Manual*. Aston Educational Enquiry Monograph No. 1. Birmingham: University of Aston in Birmingham, 1975.

FEYERABEND, P.K. Consolations for the specialist. In Lakatos, I. & Musgrave, A. (eds.). *Criticism and the growth of knowledge*. Cambridge: Cambridge University Press, 1970.

HOLTON, G. & ROLLER, D. *Foundations of modern physical science*. Reading: Addison-Wesley, 1958.

KORTH, W.W. The use of the history of science to provide student understanding of the social aspects of science. Unpublished Doctoral Thesis, Stanford University, 1970.

KUHN, T.S. *The structure of scientific revolutions*. (2nd ed.) Chicago: University of Chicago Press, 1970.

LAKATOS, I. Falsification and the methodology of scientific research programmes. In Lakatos, I. & Musgrave, A. (eds.). *Criticism and the growth of knowledge*. Cambridge: Cambridge University Press, 1970.

LUCAS, A.M. Hidden assumptions in measures of 'knowledge about science and scientists'. *Science Education*, 1975, **59** (*4*), 481-485.

MEDAWAR, P.B. Is the Scientific Paper a fraud? *The Listener*, September 12, 1963, 377-378.

PELLA, M.O. *Wisconsin inventory of scientific processes*. Madison: University of Winconsin, 1967.

POPPER, K.R. *Objective knowledge*. Oxford: Oxford University Press, 1972(a).

POPPER, K.R. *The logic of scientific discovery*. 6th impression (revised). London: Hutchinson, 1972(b).

SCHEFFLER, I. *The anatomy of inquiry* London: Routledge and Kegan Paul, 1964.

WELCH, W.W. *Welch science process inventory : Form D*. Minneapolis: University of Minnesota, 1969.

Research in Science Education, 1977, 7, 131-137.

CRITICAL THINKING IN SCIENCE – SOME PROBLEMS IN MEASUREMENT

M. J. Sullivan

Introduction

A few years ago, it seemed that no science course evaluation could be considered complete without measurements of critical thinking ability. Mostly the Watson-Glaser Critical Thinking Appraisal (WGCTA) was used, though some workers preferred the Cornell Critical Thinking Test and a few more adventurous souls constructed their own quite different instruments. To many, the subject of science instruction and critical thinking ability is now passé, although there are still unresolved disagreements as to what critical thinking is, whether or not it can be taught and, therefore, whether it can or should be measured. I now think that this is to be regretted, and suspect that in recent years a number of possibly useful ideas and techniques have been similarly "done to death" without being properly exploited.

The subject was resurrected for me when, as part of a larger research study, I asked about one hundred senior chemistry teachers in South Australia what aims they had for their students' achievement, other than a good pass in the matriculation examination, I included "the ability to think critically" in a list of commonly-held aims, but did not expect it to rank very highly when compared with such presently popular ideas as the social relevance of chemistry and favourable attitudes to science and its study. However, all but a very few of the respondents indicated that they felt this ability to be important, and that they consciously tried to provide experiences which they hoped would improve it in their students.

The intention of the study was to evaluate the matriculation chemistry course in terms of the teachers' aims, and it was clear that an instrument to measure critical thinking ability was highly desirable. The WGCTA has been the most popular of the published tests, and is still widely used in non-science areas. Its reliability is good, and its validity in its own terms is not usually questioned. However, one of the reasons for its falling popularity in the science area is criticism of its use in testing the outcomes of a course which presumably has as one of its objectives the development of critical thinking skills applicable to science-based situations. If the test does not deal with such situations, can it be validly used? Or is critical thinking a constant ability, independent of subject area?

One cannot immediately answer without being fairly clear what one means by critical thinking. The literature carries numerous definitions, some clearly different from others, and it is not the purpose of this paper to explore the possibilities. Suffice it to say that the term means something fairly definite to most people, but by no means the same thing for all.

Watson and Glaser (1964) have given a fairly complex definition, but in fact their instrument simply measures the ability defined by Ennis (1962) as critical thinking, that is, *The correct assessment of statements.* Questions requiring critical thought associated with most situations can be converted to simple statements which can be assessed in the light of evidence and/or logic. Therefore, I took "the correct assessment of statements" as my working definition of critical thinking, as Watson and Glaser obviously did.

However, the Watson and Glaser definition brings out clearly what is necessary to obtain a correct assessment of a statement, namely:

Test 1/Item 1. (E-GF)
John Smith and Ken Smith are related; they are cousins.
A. actually B. therefore C. thus D. however E. nevertheless.

2/21 (E-SC)
John Smith and Ken Smith are related; actually
A. they are cousins. B. they have different coloured hair. C. their families have been arguing with each other for years. D. they dislike each other.

9/19 (S-GF)
Light from the stars often takes many years to reach the earth. When you look at a star, you might think that you are seeing the star as it is now. , you are seeing the star as it was many years ago.
A. Simultaneously B. Accordingly C. Actually D. Similarly E. Alternatively

10/39 (S-SC)
Light from the stars often takes many years to reach the earth. When you look at a star, you might think that you are seeing the star as it is now. Actually,
A. you are seeing the star as it was many years ago. B. we can only see the stars at night because of the brightness of the sun during the day. C. the greater the distance of the star from the earth, the longer it takes for its light to reach us. D. light from the sun, our nearest star, takes eight minutes to reach the earth.

Each test contained 20 GF and 20 SC items, testing 40 different connectives; there were approximately equal numbers of S and E items in each test. Each logical connective in the list was therefore tested by at least four separate items. Connectives with various grammatical functions (e.g. *'as', 'while', 'so that'*) and connectives which occur frequently and which were known, on the basis of trial data, to be moderately difficult (e.g. *'therefore'*) were tested by means of a larger number of items. Occasionally, as many as fifteen to twenty different items were used to test particular connectives.

The sample

The target population for the study was defined as all pupils enrolled in Forms 1, 2, 3 and 4 in Victorian secondary schools during 1976. A sample size of about 150 per test per form level was planned: such a sample would lead to a maximum standard error of estimate in the item facility values of around 4%. To obtain the desired sample, a systematic sample of 46 Victorian secondary schools was selected from the ACER's sampling frame. This sampling frame is organized on the basis of school type (e.g. high school/technical school/Catholic/independent . . .) and location (metropolitan/rural, listed in postcode order). A systematic sample therefore ensures proper representation of school types and demographic regions. Occasionally, a selected school was unwilling to participate in the project, and was then replaced by the next school on the list.

Every school in the sample (except for one tiny school far from Melbourne) was personally visited by a representative of the project (in most cases, a Dip.Ed. or higher degree student). Each school received bundles containing the 25 tests in the set in an unsystematic order. Each pupil in each class received one test at random, from a bundle. A total of 16530 pupils (about 165 per test per year level) provided useable data.

Analysis

Pupils recorded their answers by making pencil marks on optically-scannable answer sheets. These were then read and analysed by the UNISCAN computer system in the Tertiary Education Research Centre of the University of New South Wales.

Findings and discussion

Item facilities

The main statistics of interest are the item facilities (percentage correct) for each item at each form level. For the purposes of the present paper, this massive collection of data has been summarised by considering only the Form 1 and 4 data, and by reporting the data as a set of *difficulty ranges*. For example, an item with 53% correct at Form 1 and 76% at Form 4 would have a difficulty range of 5-7. The difficulty range is a pair of numbers, each number being the first digit of an item facility value. (An entry of zero in the difficulty range indicates an item facility of less than 10%). The data are presented in the Appendix.

Difficult connectives

The Appendix, although a summary, still contains a formidable collection of data. A useful way of distilling this collection further is to extract a list of the most difficult connectives. Such a list, which draws only on the Form 4 data, is presented in Table 1. For the purposes of this table, a term is regarded as difficult if the mean item facility of the various items testing that connective is less than 70%; it is regarded as very difficult if the mean is below 50% and extremely difficult if less than 30%.

Inspection of this list reveals a few grouping of related connectives. Several connectives which indicate inference are difficult: *'and so', 'consequently', 'hence', 'it follows that', 'therefore'* and *'thus'*. Another group of difficult connectives are involved in generalizations: *'commonly', 'frequently', 'generally', 'in general', 'occasionally'* and *'often'*. A large group of terms signal similarities, comparisons and contrasts: *'alternatively', 'as', 'at the same time', 'conversely', 'in contrast', 'in fact', 'in turn', 'much like', 'nevertheless', 'similar to', 'similarly'* and *'unlike'*. Several apposition terms are difficult: *'for instance', 'i.e.', 'in these examples', 'namely', 'that is'* and *'viz.'*. Some students are unfamiliar with additive terms like *'again', 'also', 'further', 'furthermore', 'in addition'* and *'moreover'*.

Three connectives are extremely difficult (mean item facility less than 30%): *'conversely', 'if'* and *'moreover'*. The inclusion of *'if'* might seem surprising. In fact, its inclusion is somewhat misleading. The group of items referred to here required students to engage in correct logical reasoning e.g. to recognize that 'If Ebenezer swallows cyanide, he will die' is logically equivalent to 'If Ebenezer does not die, he cannot have swallowed cyanide'. As can be seen from the Appendix, students did not find the more usual SC and GF items testing *'if'* particularly difficult. The sharp contrast in item facilities indicates that the ability to recognize two logically equivalent statements is much more difficult than the ability to complete a sentence correctly or to fill a gap with the appropriate connective. In simple terms, using language is easier than using logic.

The *'conversely'* result is interesting. The items test the word in its strict logical

TABLE 1

Difficult logical connectives

again (furthermore)
also
alternatively
and so
apart from
as (like)
as much as
as though
AS TO
at the same time
at this/that stage
besides
by way of
clearly
commonly
consequently
CONVERSELY
ESSENTIALLY
first, second . . . finally
firstly, secondly, thirdly
for the same reason
frequently
from
from (his/this) point of view
FURTHER

furthermore
generally
hence
here
i.e.
IF (logical use)
in addition
in contrast
in fact
in general
in other words
IN PRACTICE
in terms of
in that
in these examples
in turn (in this order)
in turn (similarly)
in (. . .) ways
including
indeed
MOREOVER
much like
namely
neither/nor
nevertheless

notice that
NOW (consider that)
now (instead)
occasionally
often
on the basis of
only if (logical use)
RESPECTIVELY
say
similar to
SIMILARLY
simultaneously
so also
so far
so to speak
such that
suppose/then
THAT IS
the fact that
therefore (it follows logically)
thus (it follows logically)
unlike
VIZ.
WHERE (the meaning of the symbols being)
whereby
while (as long as)

A connective was included in this list if the mean item facility at Form 4 level, averaged across the various items testing that connective, was less than 70%. CAPITAL LETTERS indicate a very difficult term (mean item facility less than 50%). UNDERLINED CAPITALS indicate extreme difficulty (mean item facility less than 30%).

sense. (If all A's are B's, then the converse of this assertion is that all B's are A's.) To take an example used in one of the items, whenever the sun is rising, it is in the eastern part of the sky; conversely, whenever it is in the eastern part of the sky, it is rising. This strict logical use is not understood by secondary school students (or most adults, probably). The term is frequently confused with 'in contrast', or 'on the other hand'.

The language of inference

The generally high item facilities for items testing key words like *'so'* and *'therefore'* would suggest that most junior secondary school students eventually master the more familiar connectives concerned with inferential reasoning. Difficulties with terms such as *'consequently'* and *'hence'* could then be attributed to unfamiliarity, rather than difficulty with the conceptual relationships that these terms represent.

This conclusion is derived solely from those items testing these particular terms. A slightly more complex picture emerges when one inspects gap-filling items which test other connectives, and which use inference connectives as distractors. One then finds evidence of misuse of the language of inference, sometimes widespread, indicating a lack of mastery of the underlying relationship which the connective is meant to signal. Consider the following five gap filling items. The percentage of fourth-formers choosing *'therefore'* from amongst the distractors is given in brackets.

- All mammals with pouches are marsupials. (*Conversely*) all marsupials are mammals with pouches. (61% chose 'therefore'.)
- John Smith and Ken Smith are related; (*actually*) they are cousins. (45%)
- The explorers had come to the end of their food. (*Moreover*) they would soon run out of water. (35%)
- The platypus spends much of its time swimming under water. (*Frequently*) it dives to the bottom to feed on worms. (31%)
- When an iron bar is heated it increases in length. (*Similarly*) a copper bar also expands when heated. (29%)

In a sentence completion item

- Iron rusts more quickly in hot weather. Therefore
  A. iron which is painted rusts more slowly
  ... C. air and water are necessary for iron to rust
  D. ships rust more slowly in the Antarctic than in the tropics,

53% were correct, 18% chose A and 18% selected C. These two distractors are *grammatically* sound, but do not follow *logically* from the information in the stem.

These findings all point to a widespread tendency to employ *'therefore'* as a connective between two propositions even when the second proposition does not follow logically from the first. It would appear that most junior secondary school students have learned to employ 'therefore' in a grammatically correct manner, but many have not learned to avoid using the term in logically unsound ways.

Other connectives marking relationships between evidence and inference also prove difficult:

- Lemon juice, vinegar and sour milk all contain acids as shown by the fact that
  ... C. they all make litmus go red (66%)
  D. many common foods contain acids (22%)

This suggests that almost a quarter of fourth-formers interpret *'as shown by'* as if it meant 'showing that'.

Similar problems of evidence and inference are noticeable in the following items:

- The students clapped and cheered at the end of the concert; (*clearly*) they liked the performance. (The distractor 'for this reason' was chosen by 24%.)
- The metal potassium is kept in oil because it explodes when exposed to the air. For the same reason,
  ... B. the oil stops the potassium from meeting any air (38%)
  ... D. the metal sodium is kept in oil (44%)
- Some oil from car engines drops on the road, and soaks into the road surface. Rain washes this oil out and spreads it around. Hence
  ... B. oil is less dense than water and floats to the top (38%)
  C. roads are often slippery when it rains after a long dry spell. (45%)

Taken together with the item facility findings on connectives such as *'consequently'* and *'hence'*, it is clear that many students experience some difficulty with language which connects premise and conlusion, cause and effect, and evidence and inference.

The language of contrast

Another significant area of difficulty is with connectives which relate contrasting propositions e.g. *'but'*, *'although'*, *'despite'*, *'nevertheless'*, *'in contrast'* and *'on the other hand'*.

After *'and'*, *'but'* is the most frequently occurring connective in English. It is fairly easy, but it has certainly not been mastered. Eight different items tested this connective; about 10% to 20% of fourth-formers do not appear to recognise that the connective links contrasting propositions. The difficulty of the connective increases when paired with other connectives (*'but if'*, *'but also'*, *'but even'*).

Just as in the case of the inference connectives, students tend to misuse contrast connectives by applying them in inappropriate situations. Consider the following GF items; the percentage of fourth formers choosing the *'nevertheless'* distractor is given after each.

- Students should not taste acids and bases; (*indeed*) they should not taste any dangerous chemicals. (25%)
- If you want to shift a heavy stone you can either lift it or use a lever. (*By far*) the easier method is to use the lever. (21%)
- Australia has enough oil to supply its energy needs. (*At least*) it has enough for the next fifteen years. (20%)

Other contrast connectives demonstrate similar errors:

- ... Mr Stevenson was annoyed that he would be without his car for two weeks (*apart from*) the fact that he would have to pay $200 to repair it. (44% chose *'despite'*.)
- Children's heights and weights are related; (*that is*) taller children are usually heavier children. (31% chose *'although'*.)
- All living things which have only one cell belong to a group called Protozoa (*e.g.*) Protococcus (the one-celled green alga that lives on tree trunks) and yeast (. . . . .) are Protozoa. (27% selected *'however'*.)
- ... Just as the Governor-General has the power to dismiss the Prime Minister and the Federal Government (*so also*) the Governor of Victoria can sack the Premier and the State Government. (31% selected *'whereas'*.)
- Despite ________________, some people still believe it is true.
  A. the evidence for Darwin's Theory of Evolution (23%)
  ... C. the strong evidence against the theory of Lamarck (61%)

These findings suggest that roughly a quarter of the fourth-formers in the population are willing to use contrast connectives to link propositions which do not contain contrasting information.

Confusions between inferences and contrasts

The two preceding sections have been presented separately; however, there are many items in which some students select contrast connectives to link inference propositions, and vice versa. Some illustrations:

- Ionic salts dissolve in water to form solutions which conduct electricity. Sodium chloride and potassium chloride are both ionic salts. (*Consequently*) solutions of both salts will conduct electricity. (16% chose *'On the other hand'*.)
- . . . It can happen that a Senator dies during his term of office. (*In general*), the State Government then appoints . . . . (40% chose *'Despite this'*.)
- Aircraft . . . carry insect pests . . . . (*Consequently*) planes are sprayed . . . . (17% selected *'Instead'*.)
- Banks charge you for every cheque you use. (*It follows that*) it is cheaper to make payments by cash. (22% responded *'nevertheless'*.)
- Carborundum and diamond are both made of giant molecules. All substances made of giant molecules have melting points above 1000°C. Silicon is also made of giant molecules. (*Therefore*), silicon has a melting point above 1000°C. (*'However'* : 24%)
- Rye (a cereal crop) is able to withstand colder and wetter conditions than wheat. (*Thus*) it can be grown in some areas where wheat does not grow well. (*'Despite this'* : 13%)

The reverse tendency to use inference connectives in place of contrast also occurs:

- The book said that the oil in the car engine needed changing after every 10,000 km. (*Nevertheless*) John decided to change the oil after every 5000 km. (*'Therefore'* : 16%)
- Les had been looking forward to warm sunny weather for his hike. The weather turned out cold and windy; (*yet*) he decided to go as he had planned. (*'so'* : 12%; *'consequently'*: 11%)

The language of apposition

Terms which indicate apposition are *'e.g.'*, *'such as'*, *'for example'*, *'for instance'*, *'i.e.'*, *'namely'*, *'that is'* and *'viz.'*. Inspection of the item facilities indicates that the first three of these are fairly widely understood, while the last two are very difficult.

Again, an examination of items testing other connectives reveals a more complex picture, indicating that the concept of an example has not been mastered. In the following GF items, the percentage of fourth-formers selecting the *'for example'* distractor is shown.

- The Australian Postal Commission distributes mail throughout Australia by four types of transport (*viz.*) road, rail, air, and sea. (47%)
- There are many kinds of portein molecules, but they always contain atoms of four particular elements (*namely*) carbon, hydrogen, nitrogen and oxygen atoms. (34%)
- The formula of carbon dioxide is CO_2, (*that is*) a carbon dioxide molecule contains one carbon atom and two oxygen atoms. (27%)
- The Motorcycle Racing magazine was issued quarterly (*i.e.*) it came out four times a year. (24%)
- Tetra ethyl lead helps car engines to run smoothly but it also helps cause air pollution in big cities. (*Recently*), some governments have forced petrol makers to put less tetra ethyl lead in their petrol. (13%)

The data indicate widespread confusion between *'e.g.'* and other apposition connectives. The high proportion of incorrect responses shows that many students do not understand that exemplification is neither restatement nor complete enumeration of all cases.

Relationship between difficulty and frequency of occurrence

The list of logical connectives was compiled after an analysis of 2.5 million words of running text in thirty school science texts. The connective, *'and'*, was not counted, or tested. The next most frequent connective was *'but'*, occurring (as a connective) 201 times per 100,000 words. Connectives occurring less frequently than twice per 100,000 words were not included for testing purposes.

Is difficulty related to frequency of occurrence? This question can be investigated fairly easily by comparing the mean frequency of occurrence of the difficult connectives listed in Table 1 with the mean frequency of the remaining, easier connectives. (Connectives with multiple meanings, e.g. *'as', 'now', 'where'*, of which some were difficult and others were easy were omitted from consideration in making this comparison.) The finding is that the 70 'difficult' connectives occur, on average, 10.1 times per 100,000 words; for the 112 'easy' connectives, the mean frequency is 16.4.

Frequency is obviously not the sole determinant of difficult. For example, *'evidently'* and *'conversely'* appear equally infrequently, yet the latter is far more difficult. Other factors – e.g. difficulty with the concaputal relationship represented by the connective – would contribute as well.

A simple way to investigate the relationship between frequency and difficulty, holding the conceptual factor constant, is to investigate the correlation between these two variables amongst groups of connectives of similar meaning. Table 2 shows the correlations for four such groups. The values indicate that connectives which occur more frequently tend to be easier. This finding, although hardly earth-shattering, leads to the simple and testable prediction that pupils might be able to improve their comprehension of unfamiliar logical connectives if they were exposed to them more often.

TABLE 2

Correlations between frequency and mean item facility for related groups of connectives

Group	Connectives	Correlation
Additive terms	also, further, furthermore, in addition, moreover	.37
Inference terms	consequently, hence, it follows that, then, therefore, thus	.35
Generalization terms	commonly, frequently, generally, in general often, usually	.39
Apposition terms	e.g., for example, for instance, namely, such as, that is, viz.	.61

Other findings

Space does not permit a detailed presentation of other findings from the project. A few conclusions will simply be presented here; for the relevant data and detailed description, see the forthcoming report of the project (Gardner, 1977).

The steady growth of logical connective ability through the grade levels can be seen by inspecting the grand means (i.e. the mean of the 25 test means of each 40-item test) at each level. These were 23.5 at Form 1, 25.7 at Form 2, 28.4 at Form 3, and 30.6 at Form 4. Thus the typical first-former can answer over half the items presented to him; the average fourth-former can answer three-quarters of them.

The standard deviations of the tests are high, around 6 to 9, throughout the four years. Thus there are substantial numbers of pupils who perform quite poorly on these tests, even at Form 4 level.

The internal consistencies of the tests, as measured by the Cronbach alpha coefficient, are extraordinarily high, higher in fact than in any set of tests ever seen by the writer. The hundred values range from .97 to 1.00. An explanation consistent with this finding would be that there is a single, unidimensional ability, common to the various test items, which varies considerably in its extent amongst the secondary school population, and that the test items in the item bank are extremely accurate indicators of the presence or absence of this ability.

In order to study item properties, a systematic sample of forty connectives was drawn from the list, and the item facilities of the E-GF, E-SC, S-GF and S-SC items at Form 1 and Form4 were compared. This study revealed a very high correlation (around .90) between item facility at Form 1 and item facility at Form 4. In other words, the more difficult items at the earlier grade level are also the more difficult ones at the later level. The sample of items also showed no significant difference in difficulty between GF and SC items.

One might expect a difference in difficulty between S and E items; this is indeed the case, the scientific items being significantly more difficult. However, the difference in item facilities is not large (about five percent); the difference is greater at Form 1 level (6.4%) than at Form 4 (4.7%), and at Form 1, is greater for GF items (8.2%) than for SC items (4.6%).

The study also revealed significant correlations, around .3 to .7, between the item facility values of different items testing the same connective. The highest value, .70, was between the item facilities of E-GF items and the item facilities of the corresponding set of S-GF items. In other words, given two GF items both testing the same connective, their item facilities will tend to be similar even though the sentences employed in the two items are completely different. This provides powerful evidence for the existence of an underlying intellectual skill which is common to both items. The correlations between E-SC and S-SC were substantially lower, suggesting that GF items are 'purer' measures of this skill than are SC items.

Concluding note

Logical connectives are one of the principal ways in which linkage between propositions is achieved in English. As the present research has shown, the ability to answer items testing comprehension of logical connectives appears to be an extraordinarily highly internally consistent cognitive skill. By the fourth year of secondary school, most students can understand most connectives, but there are still many connectives, particularly the less frequent ones, which prove difficult.

Much research remains to be done in this area. Further basic research might include factor analytic (construct validity) studies to investigate the relationships between logical connective ability and other cognitive skills, and experimental studies to investigate how students store and process propositions related by means of logical connectives. Applied research might include the production of various diagnostic tests, drawing upon the item bank of the present project, to help teachers identify pupils with specific difficulties, and the preparation of instructional materials and techniques to increase pupils' understanding where necessary. Although such applied research could probably be done without too much difficulty, its products are unlikely to significantly affect school practice until most science teachers become convinced that language development is an important part of science education.

Acknowledgements

I am deeply grateful to many people who have helped me in this project: to Sue Taylor, for the superb way in which she took charge of virtually all the administrative aspects of the project; to Lawrie Schafe, for his help during the trials of the tests; to Eileen Badrian, who typed the tests and transcribed data; to Ian McLean of the University of New South Wales for his supervision of the final data analysis; to the many Monash students and staff members who assisted by writing and editing test items, visiting schools, and coding and sorting answer sheets; and to the principals, teachers and pupils of the schools in the sample. The project has been funded by the Education Research and Development Committee of the Australian Government.

References

GARDNER, P.L. Logical connectives in science: a preliminary report. *Research in Science Education*, 1975, 5, 161-176.

GARDNER, P.L. Logical connectives in science: some preliminary findings. *Research in Science Education*, 1976, 6, 99-108.

GARDNER, P.L. *Logical connectives in science.* Report to the Education Research and Development Committee, 1977.

NECHAMKIN, H. *The chemistry of the elements.* New York: McGraw Hill, 1968.

Appendix

Difficulty ranges

Logical connective	Everyday GF	Everyday SC	Scientific GF	Scientific SC
above all	8-9	4-7	3-5	5-7
according to	6-8	2-6	7-9	3-6
according to	7-9	3-6		
actually	2-4	8-9	7-9	4-7
after	8-9	7-9	6-9	3-5
again (=once more)	7-8	8-9	5-6	7-9
again (furthermore)	3-4	5-7	4-6	4-6
also	7-7	5-7	3-5	3-4
alternatively	3-6	3-7	3-6	3-5
although	7-9	7-8	8-9	6-8
although	6-8	6-8	5-8	5-7
and also	8-8	7-8	8-8	5-7
and so	7-8	7-8	3-6	2-5
apart from	3-4	6-7	7-7	5-7
apparently	8-9	9-9	6-6	6-9
apparently	6-8	7-9	3-6	5-8
as (=because)	8-9	8-9	5-7	4-4
as (=because)	6-8	8-9	6-8	6-9
as (=because)			5-7	6-7
as (=like)	4-6	4-6	1-4	3-5
as (=while)	9-9	9-9	8-9	5-8
as (=while)			7-8	7-9
as a result	7-8	4-7	6-8	7-8
as a result	7-9	7-9	8-9	7-8
as follows	6-9	5-8	2-5	3-6
as if	8-9	9-9	5-7	5-8
as long as	9-9	7-9	8-9	7-8
as much as	9-9	8-8	3-5	3-4
as shown by	6-6	7-9	2-5	2-6
as shown by	4-6	6-9		
as soon as	8-9	7-8	8-9	7-9
as though	4-6	6-8	3-5	3-6
as to	4-6	4-7	1-2	1-3
as we have seen	8-9	7-9	7-7	6-6
as well as	8-9	5-8	7-8	6-8
as well as			7-8	6-8
as you know	5-7	7-9	8-8	5-8
at first	7-8	6-8	7-7	5-8
at least	9-9	8-9	6-6	6-8
at the same time	7-7	4-7	7-7	3-5
at (this/that) stage	7-8	4-6	6-8	3-5
basically	6-8	6-8	6-8	3-6
because (=cause)	9-9	8-9	5-8	6-8
because (=cause)	5-8	5-8	6-8	3-6
because (=cause)			5-8	6-8
because (=reason for knowing)	8-9	6-7		
because (tautological sense)	7-8	4-6		
before	9-9	8-9	8-9	9-9
before long	8-9	7-8	8-8	5-8
besides	3-3	7-8	3-4	4-6

Logical connective	Everyday		Scientific	
	GF	SC	GF	SC
both	7-8	6-8	8-8	5-8
but	8-9	7-8	7-8	5-7
but	5-7	5-8	7-8	4-7
but also	6-6	6-8	6-6	8-9
but even	8-9	6-7	4-7	6-7
but if	3-6	6-9	6-7	4-6
but if			6-7	7-8
by	9-9	8-8	8-9	6-8
by far	6-8			
by far	5-6	5-8	4-6	6-8
by means of	7-9	6-6	7-9	8-9
by way of	2-5	4-7	3-5	4-6
certainly	2-3	8-9	7-8	5-8
clearly	2-6	8-9	2-2	2-6
clearly	2-5	7-9		
commonly	6-6	6-8	4-4	3-6
commonly			8-9	7-8
compared with	8-9	7-8	6-9	4-7
consequently	3-7	4-7	1-6	5-8
consequently			2-5	4-5
consequently			2-5	3-6
considering that	7-9	4-7	6-8	4-6
conversely	2-4			
conversely	1-3	0-1		
conversely	0-3	3-5	0-1	2-3
depending upon	7-9	6-8	4-7	4-7
despite	7-9			
despite	7-9	4-7	4-6	3-6
due to	7-9	5-8	6-9	8-9
due to	6-9	7-8	6-8	5-7
e.g.	8-9	5-8	3-5	3-5
e.g.	7-8	7-9		
either (end of sentence)	8-8	8-9		
either (. . . . or)	7-8	6-8		
either/or	6-8	6-7		
(either. . . .) or	9-9	6-9	7-8	5-6
either (. . . . or)			5-8	4-7
either/or			4-7	4-7
especially	8-9	7-8	8-9	6-8
essentially	1-3	2-5	1-2	5-7
even	8-8	8-9	6-8	5-7
even			4-7	4-7
even if	6-7	7-8	5-8	5-7
even though	7-8	6-8	5-7	4-7
even when	8-9	6-6	7-8	4-6
eventually	8-9	6-8	0-8	5-7
evidently	5-8			
evidently	5-7	6-8	3-6	4-6
exactly as	5-8	5-8	7-7	
exactly as			6-9	7-7
except	9-9	8-9	7-8	8-9
finally	8-9	7-8	8-8	6-8

Logical connective	Everyday		Scientific	
	GF	SC	GF	SC
*first, second, third, finally	3-6			
*first, second, third, finally	3-7			
*firstly, secondly, thirdly	4-6			
*firstly, secondly, thirdly	3-7			
for	8-8	7-9	8-9	6-7
for example	8-8	6-9		
for example	8-9	6-8	5-7	1-4
for instance	6-8	4-7	4-5	1-4
for the same reason			5-7	2-4
for this reason			6-9	4-7
fortunately	8-9	7-9	8-9	7-9
frequently	7-7	7-8		
frequently	2-3	5-7	3-5	5-8
from	7-8	5-7	6-6	3-5
from his point of view	4-6	4-6		
from this point of view			5-6	4-6
further	3-5	5-7	1-2	3-5
furthermore	6-8	3-5	3-6	4-6
generally	4-5	6-8	5-6	5-6
hence	1-4	6-6	2-6	4-7
hence	2-5	3-4	2-6	4-6
here	3-5	3-5	4-8	5-8
however	9-9	8-9	6-8	3-4
i.e.	2-4	4-7	3-6	6-8
if	8-9	8-9	9-9	7-8
if (. . . . then)	8-8	7-8		
(if . . . .) then	8-9	6-8		
if/then	5-6	4-7	4-6	3-5
**if (=logical use)	1-2			
**if (=logical use)	0-1			
**if (=logical use)	1-0			
if so	5-6	6-9	6-7	7-8
in addition	5-8	4-7	3-6	0-2
in addition			3-4	3-5
in case	8-9	7-9	6-7	1-3
in contrast	2-6	1-3	3-7	3-6
in fact	7-8	2-5	7-7	1-3
in general	2-2	4-7	1-2	2-5
in general	5-8	5-8		
in (many/most) instances	6-8	6-8	3-4	4-8
in order to	7-9	8-9		
in order to	7-9	7-8	8-9	8-9
in other words	7-7	3-6	3-4	4-7
in other words	6-7	6-8	5-7	7-9
in particular	5-8	7-8	4-7	6-7
in practice	1-5	3-6		
in practice	1-4	2-6	1-4	2-4
in terms of	2-4	7-9	2-4	3-6

*This item employed a unique format, in which several jumbled sentences had to be placed in correct order.

**This item employed a different format; a sentence was presented in the stem and another logically equivalent sentence had to be recognized amongst the options.

Logical connective	Everyday		Scientific	
	GF	SC	GF	SC
in that	4-6	8-9	2-5	4-6
in these examples	5-8	5-6	4-6	3-5
in this case	7-8	5-8		
in this case	6-8	6-8	3-6	4-6
in turn (=in this order)	5-6	4-6	4-6	4-5
in turn (=similarly)	4-4	3-6	7-7	6-9
in (. . . .) way(s)	6-8	8-9		
in (. . . .) way(s)	6-6	4-7	3-4	5-7
including	5-7	2-4	6-7	2-5
indeed	5-7	6-8	3-5	6-8
indeed	2-4	4-6	3-3	4-5
instead	9-8	8-9	5-7	4-8
instead			7-8	7-9
instead			6-7	6-8
instead of	9-9	8-9		
it can be seen that	6-9	6-7	6-9	5-7
it follows that	1-2	3-6	1-3	8-8
it seems clear that	8-9	8-8	7-8	7-9
it would appear that	7-9	7-9		
it would seem that	5-8	7-9	4-6	6-8
just as (=in a similar way)	6-7	5-7	9-9	8-9
just as (=in a similar way)			6-8	5-7
just as (=in a similar way)			8-9	6-7
just as (=at the same time as)	8-9	7-8	8-9	8-8
known as	6-8	6-9	6-8	5-7
moreover	0-2	1-3	0-1	3-5
moreover			0-2	2-2
much like	9-9	8-8	6-7	1-3
namely	2-6	6-8	3-5	3-7
neither	9-9	7-9	7-8	4-7
neither . . . . nor	4-6	6-7	4-5	4-8
nevertheless	5-8	4-6	4-7	3-5
next	8-9	6-9	5-7	8-9
next			7-8	6-8
nor	5-7	9-9	7-8	6-8
not only . . . . but	8-9	5-8	8-9	6-8
not only . . . . but also	8-9	7-9	6-9	5-7
not only . . . . but also			8-9	6-7
notice that	7-8	8-9	6-7	3-3
notice that			3-4	4-7
now (=at this time)	7-9	7-9	8-9	7-8
now (=at this time)			8-9	6-7
now (=consequently)	5-6	6-8		
now (=consider that)			0-0	5-8
now (=instead)	5-7	4-4		
nowadays	8-9	7-9	7-9	5-7
obviously	6-8	6-8	4-7	5-7
occasionally	5-7	5-7	5-6	5-7
of course	5-5	5-8	5-7	6-8
often	4-4	5-7	6-8	5-8
often			6-7	4-6
on the basis of	4-7	3-6		
on the basis of	2-5	7-8	4-6	5-7

Logical connective	Everyday		Scientific	
	GF	SC	GF	SC
on the other hand	7-8	3-5	7-8	6-8
once (=after)	7-9	8-9	4-7	5-7
only (=but)	5-7	8-9		
only (=but)	4-6	5-8		
only (adverb)			6-7	5-7
only if	6-7	3-6	7-9	6-9
*only if (=logical use)	3-5			
or	9-9	8-9	7-8	5-7
other than	6-8	6-8	5-7	4-7
otherwise	8-9	7-8	8-9	7-9
particularly	6-9	6-8	6-8	4-7
perhaps	9-8	8-9	7-7	6-8
plus	9-9	6-8	8-9	
possibly	7-8	7-8	7-8	6-9
probably	7-8	6-7	6-7	7-8
provided that	8-8	7-9		
provided that	5-8	5-7	5-8	5-7
rather than	8-9	8-9	4-7	6-7
really	7-7	7-9	8-8	4-6
recently	6-8	7-8	5-7	6-9
recently			5-7	
recently			7-8	5-5
respectively	2-6	1-3	1-4	3-4
say	4-5	5-8	2-3	4-6
similar to	6-6	6-7		
similarly	2-3	4-6	2-4	2-3
similarly			1-4	4-3
simply	9-9	9-9	8-9	5-8
simultaneously	3-6	2-6		
simultaneously	3-6	3-6	3-6	3-7
since	7-8	7-9	3-7	6-8
since			5-7	4-6
since			5-8	4-6
so	9-9	7-9		
so	9-9		9-9	9-9
so also	1-4	4-6	3-5	3-6
so as to	5-8	7-8	6-8	6-8
so far	5-7	4-7	4-5	4-6
so long as	7-9	6-8	5-6	6-8
so that (=with the result that)	6-8	7-8	6-9	7-8
so that (=with the purpose that)	9-9	7-9	9-9	8-9
so to speak	4-7	5-8	2-6	4-6
somehow	8-9	6-8	5-5	5-7
sometimes	8-8	6-9	7-7	8-9
still	7-8	6-7	7-9	5-8
such	6-8	4-6	7-7	5-7
such as	8-9	6-7	6-8	4-7
such that	3-5	5-8	2-4	5-7
suppose that	6-7	5-7		
suppose . . . . (then)			5-6	4-6

*This item employed a different format; a sentence was presented in the stem and another logically equivalent sentence had to be recognized amongst the options.

Logical connective	Everyday		Scientific	
	GF	SC	GF	SC
suppose . . . . (then)			3-5	3-6
(suppose) . . . . then			6-9	3-7
that is	0-3	2-3	1-5	3-7
the fact that	5-8	6-8	4-5	
then	8-9	7-8		
thereby	7-9	3-7	6-8	4-5
thereby			5-7	5-6
therefore (=it follows logically)	8-8	6-8	5-6	6-8
therefore (=it follows logically)			6-6	2-5
therefore (=for this reason)	8-8	6-9		
therefore (=for this reason)	6-7	7-8	8-9	7-8
therefore (=for this reason)	7-8	6-8	7-8	5-6
this means that	7-8	6-7	5-7	4-8
this means that			8-8	4-7
this suggests that	4-8	5-8	5-8	4-6
this suggests that			[illegible]	5-7
though	6-7	5-8	5-7	3-6
thus (=it follows logically)	3-5	4-7	3-6	4-7
thus (=for this reason)	4-6	5-8	3-7	5-8
thus (=for this reason)			3-7	4-7
thus (=thereby)			5-8	4-7
to	8-9	9-9	6-8	7-8
together with	8-9	8-9		
together with	8-9		4-5	3-5
too	9-9	8-9	9-9	5-8
unfortunately	8-9	6-8	8-8	6-8
unless	9-9	8-8		
unless	7-9	5-8	7-8	8-8
unlike	5-6	5-7	5-6	2-6
until	8-9	7-9	5-7	6-8
usually	3-5	3-5		
usually	8-8	5-8	6-8	7-8
viz.	0-1	7-8	1-2	4-7
when	8-9	7-8	8-9	7-8
whenever	9-9	8-9	8-9	8-9
where	9-9	9-9	6-8	5-7
where (=the meaning of the symbols being)			2-2	5-7
whereas	5-9	3-6	6-8	7-8
whereas	4-6	2-6		
whereby	2-6	3-3	2-7	4-6
whether	7-9	5-8		
whether	9-9	7-9	8-9	7-8
whether	7-8	5-8		
while (=but)	6-7	7-9	6-8	4-7
while (=during)	9-9	9-9	8-9	6-8
while (=as long as)			3-5	4-7
whilst	8-9	8-9	5-8	5-7
yet	4-7	5-8	6-8	6-8
yet	7-8	8-9	6-7	4-5

Research in Science Education, 1977, 7, 25-32.

A MODEL OF COGNITIVE PROCESSES

Richard T. White

Introduction

In many instances research would be improved if it were based on a rather more detailed model of what is going on in a learner's head than is normally the case. For often there is no explicit model, and the learner is treated as some form of black box. Things are done to the box, which responds in certain ways, and even when these are in tune with what the researcher expects he can say little more than "I thought it would be like that". A good example of work that lacked a detailed model was the very popular field of discovery learning in the late fifties and early sixties. Enthusiasts on both sides, pro and con discovery, had remarkably little to say about why their favoured method should be superior. How the instructional method interacted with the learner was left unstated. The disadvantage of an absent or undifferentiated model is its sterility. No hypotheses for further study are generated, and no transfer of judgement to previously unconsidered practices is possible. The discovery learning example illustrates this admirably. The undifferentiated model has a further weakness, in that rational explanations of observed effects are hardly possible. The discussion sections of articles that involve such a model often are discursive, and flit from one possible interpretation of the results to another, frequently ending with the give-away statement that "further research into this question is necessary".

A model that is explicit about internal events will generally be useful in generating new hypotheses, in judging the probable worth of new practices, and in interpreting research results. It will be more specific about individual differences than a black box model. The value of a detailed model is only to some extent, and not entirely, dependent on the accuracy of its correspondence with reality, since even discovering where it is wrong leads to an improved and more useful version. Just as the wave and particle models of light are simplified and imperfect representations of the phenomenon itself, but are still useful, so may an imperfect model of cognitive processes be useful.

The model that I wish to put forward is illustrated by figure 1. We will consider its parts in turn.

Events

Events go on around the learner. Some are dynamic and ephemeral, others static and enduring. The learner observes only a fraction of the events around him, though it would seem better to emphasise that there is not a dichotomy between observed and not observed but rather a degree of consciousness of observation. Consideration of this part of the model raises questions worth researching. What characteristics of events make them observable? What individual differences exist in observing, and how are these related to later performances? Are observing skills trainable?

Some events are deliberately arranged by the teacher (using teacher in the broadest sense, including parent, peer, or adult model, as well as functionary in an organized school system). Some of these are intended to be directly and specifically effective, as in teaching children the properties of acids and bases, while others are intended as background and gen-

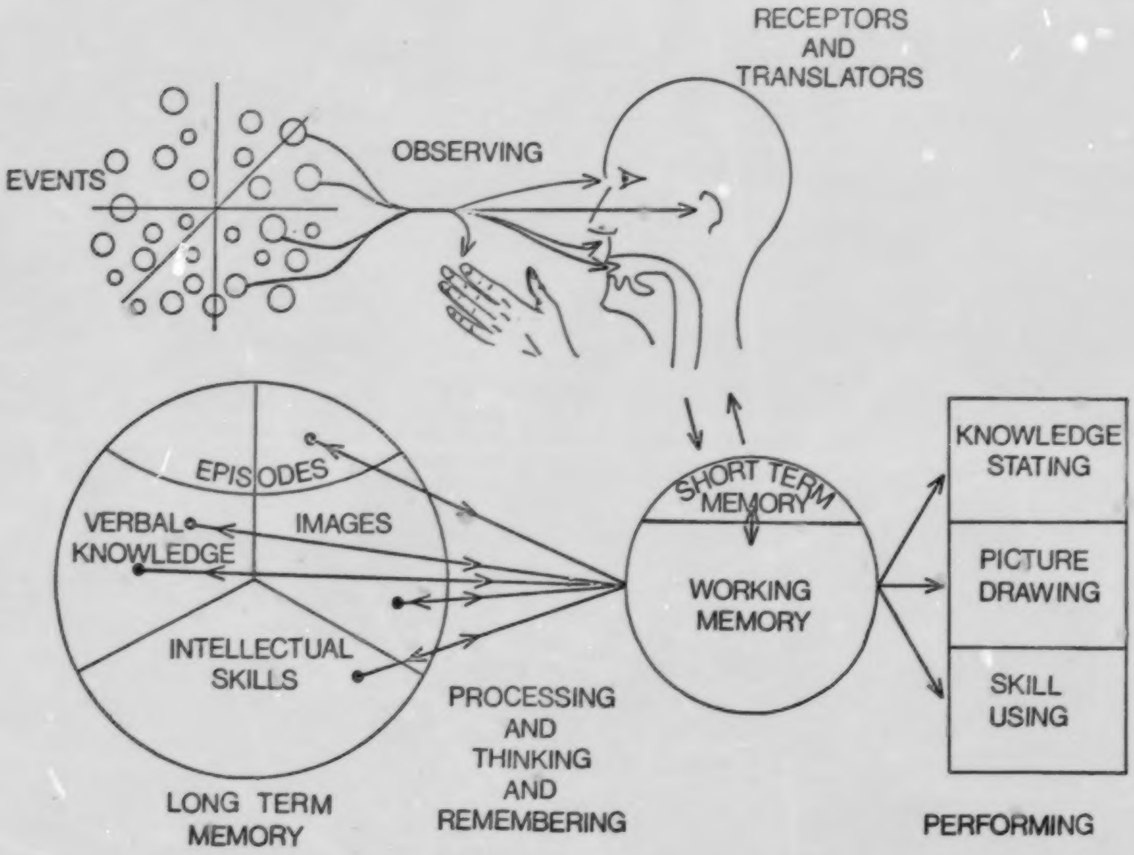

RECEPTORS
AND
TRANSLATORS
EVENTS
OBSERVING
EPISODES
VERBAL KNOWLEDGE
IMAGES
INTELLECTUAL SKILLS
SHORT TERM MEMORY
WORKING MEMORY
KNOWLEDGE STATING
PICTURE DRAWING
SKILL USING
LONG TERM MEMORY
PROCESSING
AND
THINKING
AND
REMEMBERING
PERFORMING

FIGURE 1

A model of cognitive processes

erally effective, such as arranging a science room with interesting displays, charts, and pictures. Other events are not intended to be instructive, but which a learner can use to gather information. Children can look at things around them, at puddles, caterpillars, machines, or home chemicals, and can learn from them. Of course, there are many dimensions of teaching methods, some of which are in fashion at any one time. It is perhaps interesting that through contemplating this model I have been led to specify three dimensions which I had not thought of before, and which have rarely if ever been considered in research. Two of these are mentioned above, the dimension that runs between contrived and incidental instructional events, and that between direct and enveloping events. The third is between an environment that is rich, or diverse, in events, and one that is poor, or monotonous. No doubt further contemplation would produce more dimensions. Not all of these will be relatable to other parts of the model, and therefore would be of lesser interest to me in the short term. However, in the long term they are important because they raise the question of why are they not related to the model? Should the model be extended so that these dimensions of events can be assimilated? To be fully useful, the model should be capable of predictions about the relative performances of learners subjected to different levels of a dimension of events for as many dimensions and combinations of dimensions as possible. The present model has things to say about a number of dimensions, but has its limits.

Processing

Learners translate observed stimuli into a more-or-less meaningful form, which is held in a short-term memory buffer. The process of translation is of great interest to psychologists but has been of little interest to educologists. Yet there are areas where it can be of direct importance. In early learning children acquire many skills of translation, gradually increasing their capability of making sense of their environment. In later years we take these skills for granted, but occasionally we find a new one has to be acquired. One example is learning what one is seeing in a microscope. Those of us who have acquired a translational skill of this type may easily forget how difficult it is to acquire it, and may overlook it in planning teaching or research.

Figure 1 does not bring out clearly another aspect of the model, that the processes of observing and translation are affected by the observer's cognitive structure. What attracts his attention, and what he makes of the stimuli he receives, are functions of his existing knowledge, if that term is made broad enough to include his beliefs and attitudinal set, which in turn was affected by earlier observations. The whole system then is highly interactive.

Certain properties of the short-term memory store have been studied in detail by psychologists, but since psychologists and educologists may have different purposes there may be room for further useful work. Psychologists, for instance, have concentrated on issues such as the capacity of short-term memory for arbitrary symbols, and have comparatively neglected its properties with respect to meaningful information. The limited capacity of short-term memory needs studying in education, because it means that learners can focus on only a small fraction of the events going on around them. This suggests that one individual difference of importance in education is the size of the learner's short-term memory. Those with larger stores should better be able to make use of rich environments; those with smaller stores may benefit more from close direction and control of attention.

Information held in short-term memory can be forgotten, or can be passed on to the working memory. I owe this three-stage model of memory stores, of short-term, working, and long-term memories, largely to Feigenbaum (1970). Working memory has a much greater capacity than short-term memory, and can hold information for longer periods. Some of the things you observe may never get beyond short-term memory, and are lost as soon as they are displaced by new observations. A string of nonsense syllables, for instance, or something your wife claims to have said to you, is displaced by the next, immediately following similar event. Other observations pass into working memory, and are held there for hours without necessarily being processed into long-term memory. The events of yesterday are still recallable to a large extent, but will gradually be lost unless they are processed; that is, passed into long-term memory.

Processing consists of linking new information to existing knowledge. It is a phenomenon well worth studying, since it has the potential of providing great improvements in children's learning. Rothkopf's (1970) notion of mathemagenics, the science of designing methods that will enhance learning, concerns processing. Rothkopf worked on one possible line, that of embedding questions in text. These questions improved retention even of parts of the text that the questions did not directly cover. According to the model, this improvement occurred partly because the questions maintained the focus of learners' attention on the text, but chiefly because they were encouraged to process the information in the text instead of either letting it run through short-term memory or be held temporarily in working memory. Note that the model allows us to identify four levels of attending to a text, or to a lecture. The first is where the events are not observed at all, and do not get selected into short-term memory: "Our minds are elsewhere." The reader sits with the book in front of him and later realises that he is still looking at the same words as he was some time ago. The second is where the words are translated and appear in short-term memory, and are crowded out by the next lot. When this happens, we suddenly realise that we have come to the end of the page but cannot recall a word of what we have "read", and if we re-read it, it appears totally new and unfamiliar. In a lecture, in this phase we are conscious of the lecturer talking and of his words, but cannot recall anything he said. At the third level, we are taking conscious note of the information, and are holding it in the working memory. We can recall most of it immediately afterwards, and decreasing amounts at later times, until within a day or so we can recall none of it. The fourth level is where we process it in to long-term memory and can recall it for a long period. This involves attaching it to things we already have in long-term memory.

The model allows us to interpret many of the activities of a classroom teacher as attempts to encourage processing. Teachers ask questions, partly for feedback on the success of instruction and partly to give training in making inferences, but principally to focus attention and to ensure processing. The last two should be seen as distinct; questioning happens to be a powerful technique that accomplishes both purposes. Other techniques, such as orders to pay attention, may accomplish the first but not the second.

Consideration of processing leads to important questions. Can children be trained in more effective processing skills? Can these skills be self-induced, so that more value is obtained from incidental events? Which individual differences are crucial in processing? One, for example, may be cognitive preferences (Mackay, 1975). Another is paraphrasing. Some learners have a set to paraphrase instruction, to put the instruction into their own words, which necessarily involves processing, as fast as it is fed to them. Behaviour in university lectures often diverges widely on this dimension, with some students trying to operate on level four, making sense of the information as it comes, and others at about level two with short-term memories acting as filters between ears and writing hands as they attempt to take verbatim records. Given a set to paraphrase, learners will possess the skill of paraphrasing to different extents. Is this skill innate or is it developable? Other questions concern the design of instructional events to enhance processing. Although Rothkopf's line is a beginning, there are vast areas of research in this general field that as yet are completely unexplored.

Long-term memory structures

As a result of processing, information is transferred from working memory to long-term memory. Here it will be useful to consider information as being stored in three possible forms: verbal knowledge, intellectual skills, and images.

Verbal knowledge

The vast body of information people carry in their heads is largely verbal knowledge. Hardly a day goes by without some fairly permanent accretion to it. Verbal knowledge is facts or beliefs, which are stored as propositions; e.g., there are seven days in a week, Tallahassee is the capital of Florida, $P_1V_1 = P_2V_2$, educational research is a waste of time, monocotyledons have parallel venation. There are many interesting divisions of verbal knowledge, such as that between explicable or derivable propositions and inexplicable ones, which demand more space for proper consideration than is available here. The division I mentioned above is of particular interest, because it may lie at the heart of what is meant by "understanding" of some information, and because it is an area of overlap between my own interests and those of my colleague Paul Gardner, in his work on logical connectives — explaining-type words such as *because, therefore,* and *consequently*, and qualifiers such as *but, however,* and *on the other hand* (Gardner, 1975).

The model goes as far as saying that propositions may be linked; that is, associated by the learner because he perceives either that they contain common concepts or that they contain highly related ones as far as he is concerned. For example, most science students would link the two statements:

Momentum is a vector quantity.
Energy is a scalar.

White and Gagné (1976) provide a more extensive description of linking. The model implies that linked propositions will be retained better than unlinked, and will be more readily employed in making inferences, i.e. in transfer. As well as being used in making predictions, the modal provides an opportunity to interpret Ausubel's (1968) statements about meaningful verbal learning. It implies that new information will be readily assimilated when it consists of propositions that are readily linked to extensive bodies of propositions already in the learner's long-term memory. It can explain the failure of some advance organizer experiments either through possible failure to process the advance organizer, or through presence in the learners of alternative bodies of propositions with which the new information could be linked. These explanations are not only easy to identify by consideration of the model, they also imply specific amendments to the design of advance organizer experiments. The experiments should include mechanisms to induce processing of the organizer, a check of whether propositions from it are present in the learners' memories, and an investigation of the associations actually made after assimilation of the text information.

Intellectual skills

Intellectual skills are stored memories of how to perform a class of tasks, in contrast to single facts. Gagné (1970) listed several subdivisions of skills, the most important of which in secondary school and tertiary science are concepts, rules, and cognitive strategies. Concepts are classifications; we can infer that someone has the concept of liverwort if he can identify liverworts of any shape or form. Rules are generally algorithms, knowledge of a sequence of operations: if someone can state the properties of the hydride of an element given its position in the periodic table, or can find the sum of two vectors, he has certain rules at his command. Cognitive strategies are capacities to operate on one's own thought processes; induction is an example.

Intellectual skills may be linked to one another, and also to elements of verbal knowledge. For instance, a learner may have the ability to solve gas law problems, which may remind him of another skill he has, of transforming centigrade temperatures to absolute or even to Fahrenheit. These skills may be associated by him with statements of Boyle's and Charles' Laws, to facts about the application of the gas laws in industry, to information about Boyle's life, and so on. An interesting research issue is whether these linkings enhance recall; the model implies that they do. The model also indicates that in investigating that issue, one cannot simply compare two treatments, one of which involves teaching of skill and verbal knowledge and the other the skill only, in their effect on recall of the skill, but must include a check of whether the verbal knowledge is processed and *linked to the skill.* If the researcher obtains a positive result, the check was unnecessary because the result implies that processing and linking took place. But if no difference is found in ability to recall the skill, the check is important because it can help decide between the alternative explanations that the model is wrong or that the instructional method was ineffective in providing linked verbal knowledge. Thus the model not only suggests hypotheses for study, but also influences the design of investigations so that they become more critical tests of the model.

Images

Images are pictorial or diagrammatic representations of information in long-term memory. They can be linked with other images, with verbal knowledge elements, and with skills. The image of an ammeter — black case, round, arc-shaped scale, terminals — may cause you to think of a symbol, a circle with an A in it, and of a proposition such as "Ampere was a Frenchman", and of your skill in solving Ohm's Law problems. Much teaching, particularly science teaching, is concerned with providing images. Demonstrations (which are a fertile source of images), pictures, and diagrams have two purposes, one to help initial comprehension, and the other to aid recall. It is interesting that this distinction, which seems obvious now, did not occur to me until I had formed the model and was contemplating it.

There has been a lot of research into imagery, mostly generated by Paivio (1971), but it has barely touched on the effect of pictures on retention of verbal information and intellectual skills. Holliday (1975) has completed one or two relevant studies, but there is room for many more.

Another major individual difference is suggested here by the model, that of image-forming tendency. It seems worth investigating its interaction with teaching modes that vary in their dependence on illustrations.

Episodes

Episodes are a special subset of verbal knowledge and images. They are recollections of personal involvement in an event. Of course, all learning involves personal involvement, but is not necessarily remembered in that way. One does not normally remember, for instance, the circumstances when one learned that the Battle of Hastings was in 1066. But each of us does recall doing things and experiencing events. Some of these recollections will be in the form of pictures, others as words. Often they may be recalled as words and converted into pictures at each recollection. Such re-creations of pictures may cause details to alter over time, so that one occasionally is shocked by the discrepancy between the recollection and the actual scene when one is brought back into contact with it.

Current notions of good teaching lay stress on the importance of episodes. Laboratory work, simulation games, field trips, are seen as good things. Why? Little overt rationale has been presented for them in the past, but it can be seen that the model provides one. Such experiences provide episodes which can be strongly linked, by good teaching, to desired skills and facts, thus enhancing their storage and availability. Research into the efficacy of episodes needs doing: remarkably little has been done, so the value of laboratories and field trips rests on faith rather than evidence.

Thinking

Thinking has a place in the model, though a relatively undetailed one. Thinking can be defined as re-ordering and re-associating existing propositions, and creating new ones or developing new skills either by logic or by imaginative leaps. In all cases this involves transfer from the largely unconscious long-term store to the largely conscious working memory for re-processing. The richer the collection of propositions, skills, and images, and the more complex their pattern of interlinking, the more creative the thinking potentially may be. There is more to creativity and to thinking than this, but the model does have something to say about them, something which is consistent with views of other writers, such as Koestler (1964), who have started from a very different point.

Performances

Performances occur either in response to an external stimulus, or under self-direction. The model lists three types of cognitive performance: knowledge stating, skill demonstrating, and picture drawing. These are closely related to corresponding structures in long-term memory, but it must be emphasized that other, linked structures are relevant too. For example, suppose a student has the external stimulus of a problem in kinematics, which requires him to demonstrate an intellectual skill. The words in the problem, such as acceleration, velocity, and displacement, are present in several propositions he holds in his head. These cues call the propositions to mind; the propositions are linked to images of graphs, and to episodic images of himself using kinematics trolleys and ticker timers. All of these recollections assist him to recall the skill, and enable him to demonstrate it by answering the question. Performing is thus a much more complex process than a simple one-to-one relation between recall of a single relevant element of memory and the overt act.

Summary of uses of the model

In the course of this paper I have mentioned several specific uses of the model. These uses are not restricted to this model alone, but apply to any model that contains an equivalent amount of detail. The major points of this paper are that the more detailed the model the more useful it is likely to be, and that research is likely to be more valuable when it is based on a detailed model. The more detailed a model, the more fragile it is; this is not a weakness, however, since paradoxically the strongest and most useful models are those which are easiest to prove wrong. Often fragile models have a dynamic strength, in that they are capable of extension or amendment as research results direct; but, even if it is totally refuted, a detailed model can be useful during its lifetime and at its death.

We have now seen how the model can help generate research hypotheses and identify areas of research that have been neglected. It can direct the design of investigations, and can help resolve conflicting results. It draws attention to specific individual differences. It links fields of research that otherwise may seem to be unrelated. It has things to say about practice as well as about research, by providing an explicit rationale, which is currently lacking, for what are believed to be good teaching procedures, and by suggesting new variations of teaching practice. This range of uses should make it valuable to researchers, teachers, and people engaged in teacher training.

References

AUSUBEL, D.P. *Educational psychology: A cognitive view.* New York: Holt, Rinehart & Winston, 1968.

FEIGENBAUM, E.A. Information processing and memory. In D.A. Norman, *Models of human memory.* New York: Academic Press, 1970.

GARDNER, P.L. Logical connectives in science: A preliminary report. *Research in Science Education*, 1975, 5, 161-176.

GAGNÉ, R.M. *The conditions of learning.* (2nd ed.). New York: Holt, Rinehart & Winston, 1970.

HOLLIDAY, W.G. The effects of verbal and adjunct pictorial-verbal information in science instruction. *Journal of Research in Science Teaching*, 1975, 12, 77-83.

KOESTLER, A. *The act of creation*. London: Hutchinson, 1964.

MACKAY, L.D. Cognitive preferences and achievement in physics, chemistry, science and mathematics. *Research in Science Education*, 1975, 5, 49-58.

PAIVIO, A. *Imagery and verbal processes.* New York: Holt, Rinehart & Winston, 1971.

ROTHKOPF, E.Z. The concept of mathemagenic activities. *Review of Educational Research*, 1970, 40, 325-336.

WHITE, R.T., & GAGNÉ, R.M. Retention of related and unrelated sentences. *Journal of Educational Psychology*, 1976, 68, 843-852.

Research in Science Education, 1977, 7, 33-40.

MEASURES OF MEANINGFUL LEARNING OF INTELLECTUAL SKILLS

N. C. Kellett, L. H. T. West and C. Woodruff

Introduction

Much of science learning, especially at the upper secondary and tertiary levels concerns acquiring "intellectual skills", more usually called solving "problems" – things like being able to integrate X^n, calculate the pH of a weak acid solution and so on. Sometimes the acquiring of specific intellectual skills (e.g. integrate X^n) are desired outcomes in their own right. Here the student needs to acquire the skill whether he "understands" what he is doing or not. At other times, the desired outcome is not the ability to perform a specific skill but the ability to use that skill in situations that are new to the learner. In this case the student needs to "understand" the rules or formulae that enable him to perform the learned skill so that he can apply those rules or formulae in the new situation. This outcome, which we have called "meaningful learning of intellectual skills" is an important outcome of science learning and deserves attention from science education research on that account alone.

There are other reasons, too, for an interest in this learning outcome. Learning theorists such as Gagné and White, who have been interested in the learning of intellectual skills for some time, are now turning their attention to the "understanding" of these skills. (Gagné and White, 1977). There are also a small number of empirical studies that have investigated this same outcome (Beeson, 1976; Mayer, 1975). This increase in theoretical interest makes the question of devising measures of the meaningful learning of intellectual skills an important one, and this paper describes such a method to be used in a larger study. The method has sufficient generalizability to be of interest in its own right. In addition to describing the measures, some preliminary results of an experiment using the measures are reported.

Previous attempts to measure this outcome have devised "transfer tasks" of different degree (near transfer, far transfer), and have judged the students' ability to perform each of these tasks dichotomously. In this study, transfer tasks were also used, but these were designed in a systematic manner and were not scored dichotomously. A search was made for different dimensions underlying the students' attempts at all of the transfer tasks. Three separate scales were defined – "Blind Application Index", "Flexibility" and "Application".

Systematic design of transfer tasks

The first stage in the systematic design of a series of transfer tasks must be a clear specification of the original learning task. In this study, students learned, by rule and example, a series of computational steps and finally learned to put these steps together to form the

> "50 ml. of 5×10^{-3} M $Fe(NO_3)_2$ solution is added to 200 ml. of 2.5×10^{-3} M NaOH solution.
> Will a precipitate of $Fe(OH)_2$ occur?
> (K_s for $Fe(OH)_2 = 1.6 \times 10^{-14}$)."

This sequence of computational steps, called the "standard solution", was thus the original learning task.

A "standard solution" could be learned as nothing more than a sequence of steps (e.g. first add the volumes, next divide first volume by this number etc.). Alternatively, it could be learned as a series of necessary operations (e.g. find the final concentration of each ion, etc.) which could be performed via the standard solution steps for the standard problem. We took the view that learning became increasingly meaningful as it tended toward 'operational' learning of the standard solution. (This definition would seem to concur with Ausubel's (1968) definition of meaningful learning as 'substantive and non-verbatim learning'). Appropriate transfer solutions were therefore solutions that were substantively or operationally equivalent to the standard solution, but whose computational steps were not all identical to the standard computational steps of the standard solution.

For each transfer "problem", one or more of the standard steps were modified so that they involved the same operation (e.g. find the final concentration of a particular ion) but they were computationally different. Thus the solution of these problems required the straight application of some standard steps, the modification (including omission) of other standard steps or the use of additional computational steps. For example, one transfer problem differed from the standard problem only in that the solution concentrations were expressed as g/litre, not M/litre. The solution therefore required all the standard steps and an additional step – convert g/litre to M/litre. Other solutions required more extensive adaptations of the standard solution; the problem

> "To 0.5 litre of 4×10^{-3} M $Pb(NO_3)_2$, 1×10^{-2} Moles of NH_4Cl is added. Will a precipitate of $PbCl_2$ form? (K_s given)"

involved only one liquid, and therefore several standard steps had to be omitted. An additional computational step was required to determine the molarity of NH_4Cl.

In constructing such transfer tasks it was, of course, necessary to ensure that modified and additional steps required only those computational skills for which competence had been demonstrated. This method enabled a series of tasks to be constructed, having as large a range of difficulty as desired, and also, requiring perturbation of any or all parts of a standard solution.

Measurement of performance on transfer tasks

If a student's performance on a transfer task is to be described adequately, both his overall response and the detailed construction of each solution offered would seem to be of potential interest. In the present study, the type of solution given was taken as an index of overall response. A student could give a "correct" solution, operationally defined as a solution that embodied an attempt (not necessarily successful) to provide an appropriate modification of the standard solution. This aspect of performance was measured dichotomously on each item, and the item scores were then totalled on the "correct" scale, to describe performance on the test as a whole.

A second type of solution, of particular interest, was the 'blind application' of the standard solution *in tota*. Such performance was also measured dichotomously on each item (a "1" representing a 'Blind Application'). In this context, a "non-attempted" problem could not unambiguously be coded as a 'non-blind application'; a student could have intended to apply the standard solution blindly, but been unable to do so. Allowance was made for this study by describing performance on the test as a whole in terms of the Blind Application

Index = $\frac{\text{no. of Blind Applications}}{\text{no. of attempts}}$. These two scales were not mutually exclusive in that a student could also offer a wrong solution for an item.

In considering the detailed construction of transfer solutions, it seemed unreasonable to assume that meaningful learning of the standard solution would necessarily affect the ability to reformulate and adapt parts of the standard solution and the ability to apply standard steps in the same way. Therefore, separate measures of these two outcomes were made. Each step in a solution was assigned a score (1 for each standard step; the mark for each non-standard or "flexible" step was determined by the number of procedures or computations it entailed). A student's scores for all the standard steps were totalled to give a measure of his performance on the Application Scale. His scores on the non-standard steps were totalled to provide another measure of performance on the Flexibility Scale.

Details of the four scales defined are summarized in Table 1.

TABLE 1

The four measurement scales proposed

SCALE	MEASURED QUANTITY	RANGE
CORRECT	TYPE OF SOLUTION	0-8
BLIND APPLICATION INDEX	TYPE OF SOLUTION	0-1
FLEXIBILITY	PERFORMANCE ON NON-STANDARD STEPS	0-37
APPLICATION	PERFORMANCE ON STANDARD STEPS	0-48

Characteristics of the scales

Ideally a new measuring instrument should be administered to a group of students outside the experiment to establish its characteristics. There are a number of problems with the use of this approch for a transfer test. A task is only a transfer task if the student has not previously met that task. Thus the test is valid only in relationship to the respondent's previous instruction. Further, in this case, the scoring procedures used rely on the students' using a particular solution rule (taught to them in the instructional materials) rather than one of a number of alternative solution rules that could be used. Therefore, it was necessary for the test to be trialled with students who had learned the basic skill through the learning programme to be used in the experiment. It was also desirable for the trial sample to possess a range of abilities to transfer. The procedure used was to give the learning programmes to a group of students who had completed learning of relevant background learning (the principles of chemical equilibrium) and then to administer the transfer test. The previous learning of relevant background should enable many of the students to "understand" the skill they learned in the programme. Sixty-eight eleventh grade students in one school were used for this trial.

Initially the responses were scored into the four scales described above. The uniqueness of these scales was tested by examining their intercorrelations (Table 2).

TABLE 2

Scale intercorrelations

	B.I.	FLEXIBILITY	APPLICATION
CORRECT	-.82	.96	.68
BLIND APPLICATION INDEX (B.I.)		-.85	-.44
FLEXIBILITY			.65

It is clear that the Flexibility scale and the Number Correct are almost identical. This may also be true of the Blind Application Index. However it was decided to retain the Flexibility scale and the Blind application scale as two separate measures. Clearly, the Application scale measures something different. It is interesting to note the very high correlation between the number correct and Flexibility score. It seems that the ability to correctly apply a skill to a transfer task is very dependent upon the ability to perform the steps that we have described as "flexibility steps".

Table 3 gives the internal consistencies (measured by Cronbach's alpha), the means and the standard deviations of the three scales. Determination of the internal consistency of the Blind application scale posed an interesting problem. A blind application in a particular question is scored as 1, but a failure to blindly apply (scored as zero) means a number of things – the question is correctly answered, a wrong solution is attempted or the question was omitted. Clearly, the last of these cannot be considered as a non-blind application; nor is it equivalent to the other options. As previously explained, this difficulty was overcome for the scale by calculating an index, but this did not overcome the problem of what to use for V_i, the variance of each of the items and V_t the variance of the Scale, in computing Cronbach's alpha. An estimated value was used – the rationale for this estimate is given in an appendix to this paper.

TABLE 3

Scale characteristics

	MEAN	S.D.	CRONBACH'S α
FLEXIBILITY	12.22	8.27	.83
APPLICATION	33.38	8.87	.90
BLIND APPLICATION INDEX	0.62	0.27	.73

Results

The results presented in this section were taken from an experiment in which a group of 90 students, all naive with respect to background chemical information, were required to learn the standard solution in the same way. The group was randomly split into two halves, the experimental group receiving Ausubelian-based material which was designed to promote more meaningful learning of the standard solution, in place of the placebo materials offered to the control group.

TABLE 4

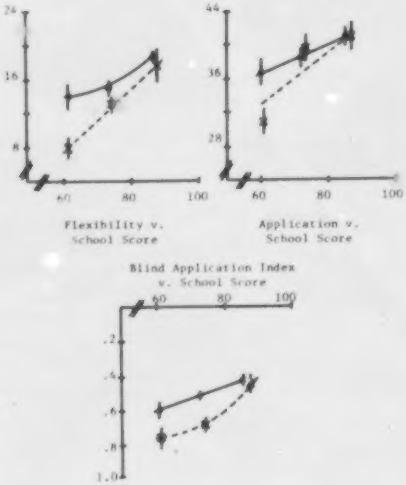
Experimental results: means and standard errors

	CONTROL (N = 45)		EXPERIMENTAL (N = 45)		t
APPLICATION	36.9	(1.3)	38.2	(1.2)	0.76
FLEXIBILITY	12.8	(1.1)	15.7	(0.9)	1.98*
BLIND APPLICATION INDEX	0.64	(0.04)	0.51	(0.04)	2.54**

*$p < .05$, **$p < .01$ (for 1 tailed t tests.)

Table 4 gives the mean and standard error for each group or the three scales considered. These results indicate a significant improvement in performance in the Flexibility and Blind Application scales for the Experimental Group but no difference between groups on the Application scale.

To investigate the interaction of Chemistry Learning Ability and Treatment, the two groups were split into High, Medium and Low sections on the basis of School Chemistry marks. The means and standard errors of these sections are shown graphically in Fig. 1.



Flexibility v. School Score

Application v. School Score

Blind Application Index v. School Score

FIGURE 1

High, medium and low section performance. o experimental x control

The Flexibility and Blind Application curves show increased divergence as one moves toward the low sections, and there is also the suggestion of a significant difference between the two low sections on the Application Scale. These results concur with the results of many Ausubelian studies. It is interesting to note the similarity of the Flexibility curves and the Blind Application curves in the light of the fairly high interscale correlation reported earlier.

Conclusion

The results reported in the previous section indicated that the Flexibility and Application scales measured different aspects or dimensions of performance on transfer tasks; these results have suggested that meaningful learning may be particularly related to the flexible dimension of performance. Although the Blind Application Index may not measure a separate aspect of transfer performance, it is nonetheless a useful scale, as it gives a succinct description of an interesting aspect of behaviour — the "non-meaningful" learning of a solution rule.

References

AUSUBEL, D.P. *Educational psychology: a cognitive view.* New York: Holt, Rinehart and Winston, 1968.

BEESON, G.W. Effects of context on hierarchical learning. Unpublished Ph.D. Thesis, Monash University, 1976.

GAGNE, R.M. and WHITE, R.T. Memory structures and learning outcomes. Unpublished document, 1977.

MAYER, R.E. Information processing variables in learning to solve problems. *Review of Education Research*, 1975, 45, 525-542.

Appendix. Estimation of Cronbach's alpha for the Blind Application Index

As stated in the text of this paper, a zero score can mean either that the student did not blindly apply the rule, or that he omitted the question. In this situation it is not valid for a non-attempt to contribute as zero to the overall scale. A method of estimating alpha, using a correction to allow for such non-attempted questions is given below. This method should be applicable to any scale in which a non-attempt cannot be scored as zero.

$$\alpha = \frac{n}{n-1}\left(1 - \frac{\Sigma V_i}{V_t}\right) \quad (n = \text{no. of items})$$

Consider first a set of n items, all attempted. If the jth person's item score is X_{ij} and his total scale score is Y_j, the covariance of the ith item scores with the total scale scores is

$$C_i = \frac{\sum_j (X_{ij} - \bar{X}_i)(Y_j - \bar{Y})}{N-1} \quad (N = \text{no. of persons})$$

The total covariance of all items with the total score is

$$\begin{aligned} C_t &= \sum_i C_i \\ &= \frac{1}{N-1} \sum_i \sum_j (X_{ij} - \bar{X}_i)(Y_j - \bar{Y}) \\ &= \frac{1}{N-1} \sum_j [(Y_j - \bar{Y}) \sum_i (X_{ij} - \bar{X}_i)] \\ &= \frac{1}{N-1} \sum_j [(Y_j - \bar{Y})(\sum_i X_{ij} - \sum_i \bar{X}_i)] \\ &= \frac{1}{N-1} \sum_j [(Y_j - \bar{Y})(Y_j - \bar{Y})] \\ &= V_t \end{aligned}$$

where V_t is the variance of the total scale scores.

Therefore,

$$\alpha = \frac{n}{n-1}\left(1 - \frac{\Sigma V_i}{C_t}\right)$$

In computational form

$$C_i = \frac{1}{N-1}\left(\sum_j X_{ij}Y_j - \sum_j X_{ij} \sum_j Y_j/N\right)$$

$$C_t = \frac{1}{N-1}\left(\sum_i \sum_j X_{ij} Y_{ij} - \sum_i \sum_j X_{ij} \cdot \Sigma Y_j/N\right)$$

The first term is not affected by non-attempting. However the second term contains $\sum_j Y_j$ which is normally a constant for all items. Because of non-attempts, the total score summation for the ith item will involve only those Y values corresponding to attempts on the ith items; therefore we should use $\sum_j Y_{ij}$ instead of $\sum_j Y_j$. We can estimate this non-attempted correction by the approximation

$$\Sigma Y_{ij} = \Sigma Y_j - n_{ai} Y,$$

where n_{ai} is the number of non-attempts for the ith item. Thus the corrected term

$$\sum_i \sum_j X_{ij} \sum_j Y_{ij}/N \text{ becomes } \sum_i \sum_j X_{ij} \; (\Sigma\, Y_j - n_{ai}\, \bar{Y})/N$$

$$= \sum_i X_i (\Sigma\, Y_j - n_{ai}\, \bar{Y})/N$$

where $X_i = \sum_j X_{ij}$

$$= (\Sigma Y)^2/N - \sum_i X_i\, n_{ai}\, \bar{Y}/N.$$

$$\text{Since } \sum_i \sum_j X_{ij}\, Y_j = \sum_j Y_j \sum_i X_{ij} = \sum_j v_j\, Y_j = \sum_i Y_j^{\,2}$$

the corrected total covariance becomes

$$C_t^{\,\text{est.}} = \frac{1}{N-1}\left(\Sigma Y^2 - \frac{(\Sigma Y)^2}{N} + \frac{\sum_i X_i\, n_{ai}\, \bar{Y}}{N}\right)$$

$$= V_t + \frac{1}{N(N-1)} \sum_i X_i\, n_{ai}\, \bar{Y}$$

and

$$\alpha^{\text{est.}} = \frac{n}{n-1}\left(1 - \frac{\Sigma V_i}{C_t^{\,\text{est.}}}\right)$$

Research in Science Education, 1977, 7, 41-49.

PRIMARY SCHOOL PUPILS' ABILITY TO SEE SCIENTIFIC PROBLEMS IN EVERYDAY PHENOMENA

David J. Symington

Introduction

Much has been said about the natural curiosity of young children and many of the science programmes for primary schools developed during the past two decades have taken this alleged curiosity into account.

This is probably seen most clearly in the British Nuffield Junior Science and its successor Science 5/13 and the Victorian Primary Science Course. These courses are built on the assumption that children are curious and will see problems in their environment which they will want to investigate. Teachers are expected to provide positive reinforcement to this natural desire to pose questions; to encourage children to believe that school is a place where pupils can ask questions and conduct their own inquiry to answer these questions. In the introduction to Chapter 1 of *With Objectives in Mind*, the guide book to the Science 5/13 project, teachers are told: 'In general the children work best when trying to find answers to problems that they have themselves chosen to investigate. These problems are best drawn from their own environment and tackled largely by practical investigations.' (p.4).

Science in the primary school is presented in that course as providing the opportunity for pupils to investigate problems in the environment. Further it is assumed that the pupils themselves will be a source of these problems. Science 5/13 teachers are advised that the authors' convictions are that 'we must help (pupils) to ask their own questions and find their own answers by first hand investigation as far as that may be.' (*With Objectives in Mind*, p.5).

It is suggested that the problems for study will arise from observation of everyday phenomena. For example, in Curriculum Guide C to the Victorian course, the topic Comparison of Liquids is introduced by advising teachers that 'This topic might arise naturally from observation of raindrops on a window, or of milk, ink, or water spilt on the floor or a desk.' (p.9). Some teachers however believe that the majority of pupils are unable to see investigable scientific problems in everyday phenomena and hence that it is not appropriate to develop a course around problems seen by pupils.

One of my colleagues tells a story which sounds apocryphal but which he claims to be true. When the Victorian course was first introduced, one of the teachers at the school where he was teaching began his first science lesson by telling the students to get out and find things to discover. He ushered the bewildered pupils out into the yard and settled back to await the inevitable result which would clearly establish, to his satisfaction, that this ludicrous course could never work. The story illustrates an extreme case of misunderstanding of the intentions of the course. However similar views are common enough to be taken seriously and have been taken in that way by Crossland (1972) in his study of Nuffield Junior Science. Crossland (1972) prepared a report on the success of Nuffield Junior Science. To gather the data he visited trial schools and held discussions with one hundred heads, assistant teachers and LEA contacts. Opinions of a further two hundred head teachers and assistants taking part in the trial were collected by postal questionnaire.

Amongst the views expressed Crossland recorded the following comments of teachers:

> 'A very small minority asked questions which could be answered by experiment.'
>
> 'Those interested in the work asked questions related to it — the dull ones accept their environment.' (p.633)

These two comments suggest two factors which could limit the teacher in using pupil posed problems as the basis for science activities. The first is that not all questions asked by pupils are suitable starting points for investigation. The second is that pupils differ in their ability to see scientific problems in everyday phenomena.

This study therefore, is designed to investigate the extent to which the questions asked by pupils form suitable starting points for investigations, and the ability of pupils to see scientific problems in everyday phenomena.

It needs to be noted that this study cannot establish whether the ability of pupils to see scientific problems in everyday phenomena is sufficient for the Science 5/13 or the Victorian Primary Science Course to be effective. However to simply begin to measure that ability in a number of contexts will provide some data for curriculum planners to make better informed judgements.

Measuring the ability of pupils to see scientific problems in everyday phenomena

Pupils meet phenomena in various ways. For example, a child may see a snail on the footpath as he walks to school or he may see a photograph of a snail in a book. A child may see another child on a swing or may see a photograph of a pendulum. In all of these situations it may be possible for a pupil to see scientific problems. So in this study children were asked about the problems they could see both when they saw photographs of everyday phenomena and when they could actually handle and play with materials.

The wording of the question asked could be an important factor in the response obtained from pupils. Many questions were used during preliminary investigations leading to the use of two forms of wording in this study. Firstly, when shown photographs of natural phenomena, the pupil was asked to suggest all the things that he did not know but a scientist might know about what he could see in the picture. Secondly, when using equipment the pupil was asked to suggest all the questions about the phenomena to which he and a group of fellow pupils could find an answer by using available materials and with the aid of their teacher. This second form of questioning tended to restrict the response of those problems which could possibly be solved by using the materials available.

In all of these different situations each separate idea was counted as a 'scientific problem'. Hence a total number of 'scientific problems' was calculated. The responses were also divided into 3 groups based on criteria established as a result of trials. These three groups were:

(i) Data-gathering problems

e.g. How fast does a snail move?
How many veins are there in the leaf?

Grade 5 and 6 pupils working with available materials, and with the aid of their teacher, could arrive at meaningful answers to these scientific problems by use of the processes of observing and measuring alone.

(ii) Data-processing problems

e.g. Will drops of water on a leaf soak into the leaf?
Why do drops of water form on the outside of a glass of water?

Grade 5 and 6 pupils working with available materials and with the aid ot their teacher, could arrive at meaningful answers to these scientific problems but the solution would require scientific processes in addition to those of observing and measuring.

(iii) Non-investigable problems

e.g. Why is the leaf that shape?
How is the glass made?

Grade 5 and 6 pupils could not obtain meaningful answers to these scientific problems through working with available materials and with the aid of their teacher.

Reliability and validity of the judge's categorization

To measure an ability such as the pupils' ability to see scientific problems in everyday phenomena one must use open-ended methods such as that proposed. It is obviously important then to provide measures of the reliability and validity of the categorizing of the pupils' responses.

The Initial data for the study was gathered by individually interviewing 59 grade 6 pupils. Each pupil was shown a set of 5 photographs and asked to suggest things that he did not know but a scientist might know. A total of 473 problems were suggested. Each of the problems was placed in one of the three categories defined. Some days later each of the problems was again categorized. Comparison of the categorization showed that 92.6% of the problems were placed in the same category on the two occasions. This figure suggests that the responses and categories are such that reliable categorization is possible. To validate the categorization two science educators with considerable primary school teaching experience were asked to categorize every tenth problem suggested by the pupils. Both judges agreed with the category suggested for 44 (92%) of the 48 problems listed. This result gave confidence to proceed.

Scientific problems seen by pupils in photographs of everyday phenomena

When the problems suggested by the pupils were categorized the total number of problems in each category for each photograph was obtained. The results are set out in Table 1.

TABLE 1

Numbers of problems in each category suggested by 59 grade 6 pupils

Category of Problems	Photograph					
	1	2	3	4	5	Totals
1. Data gathering	39	14	11	15	8	87
2. Data processing	45	14	15	10	38	122
3. Non-investigable	44	35	99	33	53	264
Totals	128	63	125	58	99	473

Key to Photographs = 1 – snail leaving distinctive 'silver trail'.
2 – small girl bouncing ball.
3 – several drops of water on a leaf.
4 – child on a moving swing.
5 – glass of water with condensation on the outside of the glass.

The data have been handled in such a way as to enable consideration of the two matters raised by Crossland's report.

Firstly teachers interviewed by Crossland suggested that very few of the questions asked by pupils were answerable by experiment. Table 2 combines categories 1 and 2 into a category labelled Investigable Problems. An inspection of the data in Table 2 shows that overall more than half of the problems suggested by these pupils were classified as non-investigable.

TABLE 2

Number of investigable and non-investigable problems suggested by 60 grade 6 pupils

Category of Problems	Photograph					
	1	2	3	4	5	Totals
Investigable	84	28	26	25	46	209
Non-investigable	44	35	99	33	53	264
Total	128	63	125	58	99	473

The complaint of teachers that many of the questions pupils asked were not answarable through classroom investigation can be understood when one looks at these data. However, the data also suggest that some of the photographs stimulated more investigable problems in the minds of pupils than others. For example, 66% of the problems about snails were categorized as investigable compared with 21% of the problems arising from the picture of the water on the leaf. Or taking specific problems raised, the most commonly raised questions about the snail were investigable problems about the 'silvery trail'. On the other hand the most common question arising from the picture of the leaf was 'Why is the leaf that shape?' which was considered a non-investigable problem. To establish the variability across photographs a χ^2 text was applied to the data in Table 1 yielding a value of 66.7 ($p < .001$).

Thus, although the data clearly establish that many of the problems raised by pupils are not suitable starting points for classroom investigation, the variation across photographs suggests that the subject matter may be a significant factor. It appears likely that pupils can suggest more investigable problems about some phenomena than others. This matter is to be investigated more fully in a later stage of the total project.

The second teacher complaint recorded by Crossland was that there was considerable variation in the extent to which pupils suggest problems. This is likely to be a compound factor in the class room, possibly reflecting both ability and motivation. As a means of providing some evidence related to this comment the data was further analysed to look at variation between pupils in the number of investigable problems suggested.

The procedure used was one way analysis of variance with repeated measures (Winer, 1971, p.268). The results obtained are presented in Table 3.

TABLE 3

Analysis of variance with repeated measures — number of investigable problems suggested by pupils in response to 5 photographs

	SS	df	MS	F
Between people	76.9	58	1.33	3.32*
Within people	151.6	236		
Photographs	57.2	4	14.3	35.75*
Residual	94.4	232	0.40	

*$p < .001$.

It will be noted that this analysis of the data reveals significant variation both in the response to the various photographs and between the pupils in their ability to suggest investigable problems.

Reasons why pupils differ in their response to a particular phenomenon

Work reported by Dietz (1975) provided a starting point for considering the response of pupils to the phenomena. Her research suggested that pupils could be placed in three categories when confronted with particular 'problem situations'. Firstly there are those who see the problem. Those unable to see the problem Dietz sees falling into two groups. In one group are those who do not see a problem because they understand the 'problem situation'. In the other group are those who do not understand but are still not aware of the potential problem in the situation.

With this framework in mind the data collected in this first part of the study was considered and a model, shown in Figure 1 developed to explain the varied response of the pupils to the stimuli provided. It can be seen that the categories (A), (E) and (F) basically match those described by Dietz and that these categories have been further subdivided in this model. Consideration of this model has implications both for further research into the topic and for teaching. In this paper only research implications will be pursued.

If the problem-solving method suggested in the Victorian and Science 5/13 courses is a valuable learning method then it would be worthwhile to investigate ways in which the number of pupils who do see investigable problems in everyday phenomena is maximized. The data already reported suggest that a judicious choice of subject matter as the basis for this method of learning may be important. However, when this choice has been made the model proposed suggests other ways in which the number of pupils who can see investigable problems in the phenomena can be increased.

Firstly, the existence of category K suggests that a period of unstructured observation and experimentation with materials may stimulate an awareness of problems. Further to this, the nature of the materials available may be a factor in the extent to which this period of 'play' is a stimulus to problem awareness.

Secondly, it could be imagined that following a period of 'play' with materials a class discussion of observations and hypotheses could result in pupils previously unaware of problems (category L) becoming aware of a problem and pupils who previously believed they understood the phenomenon (category J) becoming aware of other hypotheses. Class discussion therefore could be important in increasing the number of pupils who see problems in the phenomena. Thus a number of research questions arise from the model described (Figure 1).

> Does an opportunity for unstructured observation and experimentation with materials help pupils to see investigable scientific problems?
>
> Does the nature of the materials given to pupils for a period of unstructured observation and experimentation influence their ability to see investigable scientific problems?
>
> Is the teacher able to use class discussion of a period of unstructured observation and experimentation to help pupils see investigable scientific problems in everyday phenomena?

Investigations into these questions have begun and the first of these will be briefly described in this paper.

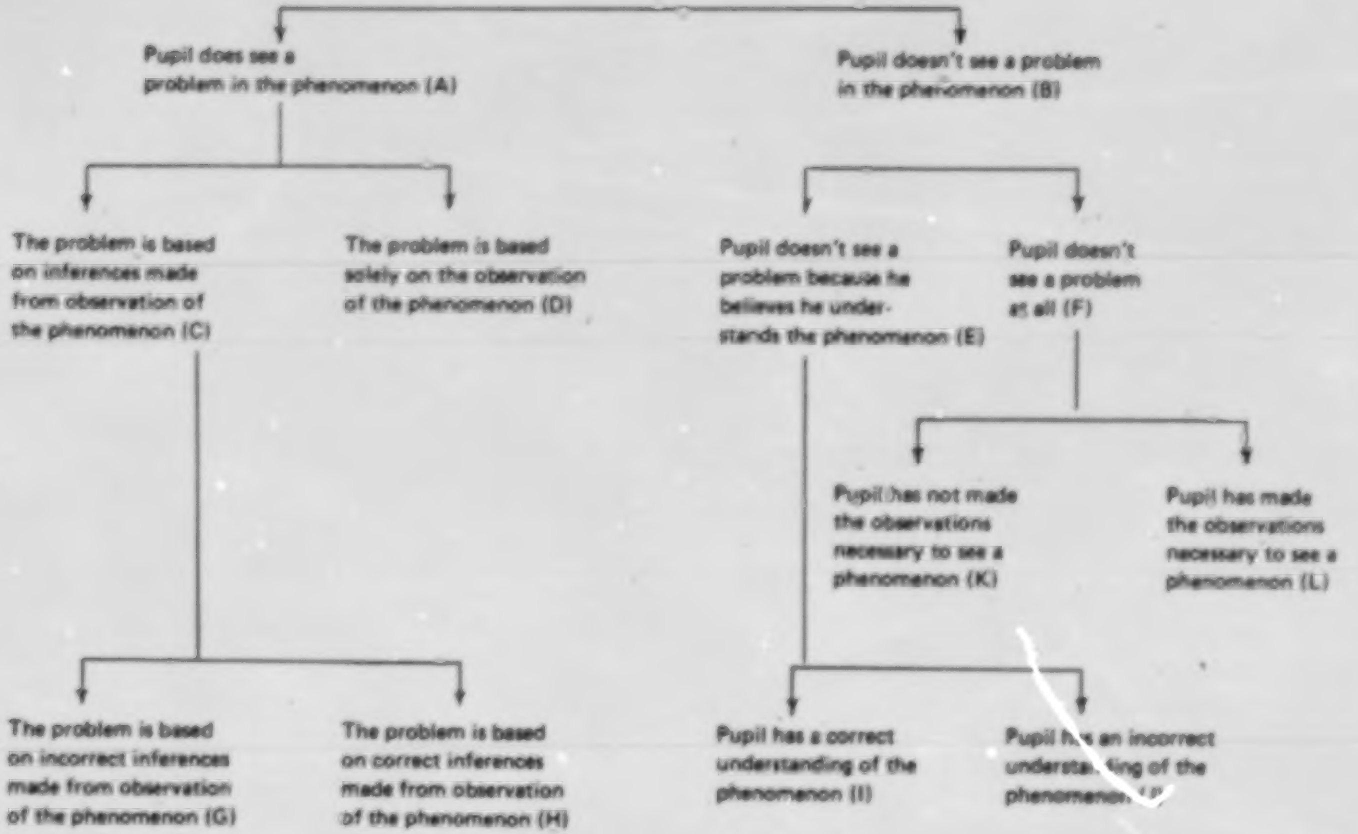

Pupil does see a problem in the phenomenon (A)
Pupil doesn't see a problem in the phenomenon (B)
The problem is based on inferences made from observation of the phenomenon (C)
The problem is based solely on the observation of the phenomenon (D)
Pupil doesn't see a problem because he believes he understands the phenomenon (E)
Pupil doesn't see a problem at all (F)
Pupil has not made the observations necessary to see a phenomenon (K)
Pupil has made the observations necessary to see a phenomenon (L)
The problem is based on incorrect inferences made from observation of the phenomenon (G)
The problem is based on correct inferences made from observation of the phenomenon (H)
Pupil has a correct understanding of the phenomenon (I)
Pupil has an incorrect understanding of the phenomenon (J)

FIGURE 1

Pupil response to an everyday phenomenon

Unstructured observation and experimentation

The first of the research questions suggested may be subsumed in the second. If it could be established that different sets of materials provided for a period of unstructured observation and experimentation have a differential effect on pupil awareness of scientific problems then clearly the period of unstructured observation and experimentation can have a significant effect upon the pupil variable.

For this investigation classes of grade 5 or 6 pupils were selected and the pupils in each class divided randomly in half. In each half class the pupils were given a ten minute period of unstructured observation and experimentation in groups and then each pupil was asked to record problems which they and a group of peers could investigate using materials shown to them by the researcher.

All of the pupils in one half of each class were given the same set of materials but these were different from the set of materials given to pupils in the other half of the same class for the period of unstructured observation and experimentation. All groups were shown the same set of materials when asked to record problems for investigation. Using Campbell and Stanley (1963) notation the experiments could be represented as

$$RX_1O$$
$$RX_2O$$

where X_1 represents a period of unstructured observation and experimentation with one set of materials, and X_2 represents a similar period of unstructured observation and experimentation with the other set of materials.

The materials provided to the pupils in the two halves were chosen in an effort to allow one half of the class to make comparisons whilst this was not possible in the other half of the class. For the first experiment groups of pupils were provided with a pendulum frame as shown in Figure 2. For one half of the class a single pendulum was hung from the frame whilst in the other half two pendulums were hung, one having a heavier bob and longer string than the other.

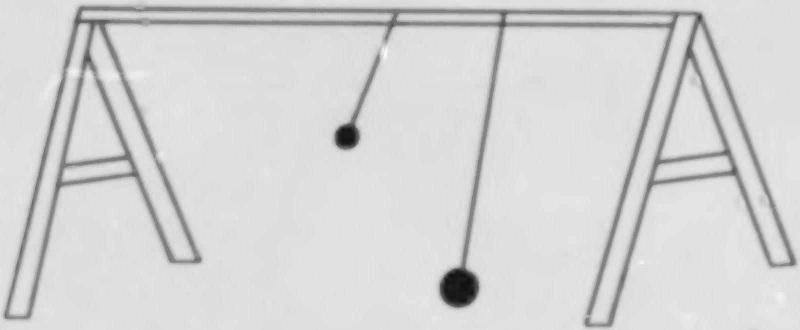


FIGURE 2

Pendulum frame used by pupils

For the second experiment each pupil in one half of the class was given a magnet and a set of objects which were all attracted to the magnet. The pupils in the other half of the class were given a magnet and a set of objects only some of which were attracted to the magnet.

It was hypothesized that pupils able to observe contrasting behaviour (that is, those with the two pendula and those with both materials which are attracted to the magnet and those which are not) would record more investigable problems than the pupils with the other materials. The results of the first experiment are recorded in Table 4.

TABLE 4

Number of investigable problems seen by pupils after a period of unstructured observation and investigation

	Pupils with two pendulums on frame	Pupils with one pendulum on frame
Number of pupils	46	37
Average number of investigable problems	3.76	2.57
Standard deviation	1.66	2.02

't' value = 3.00 ($p < .005$)

The 't' value of 3.00 indicates that there is a very low probability of the recorded differences being due to chance variation. It is reasonable to assume that providing students with the 'contrasting' pendulums will stimulate more investigable scientific problems than the provision of a single pendulum.

The results of the second experiment using magnets are recorded in Table 5. In this experiment there was not a significant variation between the two groups of pupils. Reading of the problems proposed by pupils suggested that most pupils had significant previous experience with magnetism (pupils in two classes in fact reported in-class activities) and so, upon reflection, it is not surprising that there was no significant difference between the groups in this experiment.

TABLE 5

Number of investigable problems seen by pupils after a period of unstructured observation and investigation

	Pupils with materials some of which are attracted to a magnet	Pupils with materials all of which are attracted to a magnet
Number of pupils	61	57
Average number of investigable problems	2.85	2.77
Standard deviation	1.62	1.41

't' value = 0.29 (*n.s*)

Together the results of these two experiments suggest that the materials provided for the period of unstructured observation and experimentation may have a significant impact on the problems seen by pupils if the phenomenon is not particularly familiar to the pupils. This matter needs to be, and will be, pursued in further experiments.

Conclusion

Although the experiments reported here are only a beginning to the research needed, the results obtained so far suggest that some teachers and some curriculum planners have overlooked factors in their consideration of problem solving as a learning method in primary school science.

There appear to be teachers who have rejected this approach prematurely because many of the problems children suggest when they are first introduced to this method of working are not sound starting points for investigation.

Curriculum palnners, on the other hand, do not appear to have given serious consideration to the fact that some subject matter provides a better starting point for pupil problem solving than others. Further, there has been inadequate information available for teachers on the type of classroom situations and teacher behaviour which will maximize pupils' ability to see investigable science problems in everyday phenomena. The type of research reported here will be continued and expanded to provide a clearer picture of contexts in which the ability of pupils to see investigable scientific problems is maximized.

References

CAMPBELL, D. and STANLEY, J. *Experimental and quasi - experimental designs for research.* Chicago: Rand McNally, 1963.

CROSSLAND, R.W. An individual study of the Nuffield Foundation Primary Science Project. *School Science Review*, 1972, 53, 628-638.

DIETZ, M.A. Problem solving ability and its place in primary science. *Probe*, 1975, 4, 3-10.

SCIENCE 5/13. *With objectives in mind.* London: Macdonald Educational, 1972.

VICTORIAN EDUCATION DEPARTMENT. *Science: Course of study.* Melbourne: Victorian Government Printer, 1970.

VICTORIAN EDUCATION DEPARTMENT. *Curriculum guide primary science C. Branching out.* Melbourne: Victorian Government Printer, 1969.

WINER, B.J. *Statistical principles in experimental design.* Tokyo: McGraw Hill Kogakusha, 1971.

Research in Science Education, 1977, 7, 51-60.

AN EVALUATION OF THE IMPLEMENTATION OF THE JUNIOR SECONDARY SCIENCE SYLLABUS IN N.S.W.

Sue McBurney

Introduction

In recent years there has been a trend in Syllabus development towards broad statements of aims with minimal specification of content by a central authority. Associated with this trend has been a change in the roles and responsibilities expected of teachers. Specifically, teachers are being increasingly called upon not just to follow a curriculum but to devise that curriculum themselves. Syllabus authors bequeathing such responsibility to teachers frequently do so in the belief that teachers have sufficient professional competence to design their own curricula within broad guidelines and that the experience of doing so further enhances their professional development.

The N.S.W. Junior Secondary Science Syllabus is part of this trend to local development of curricula. It consists of a one-page statement of aims (16 aims in all), followed by a brief explanation of each aim, and finally a one-page statement of content accompanied by a rationale for the form of the statement of content. The Syllabus is designed for years 7 to 10 (ages 12 to 15) for all ability levels. As such, it is a far cry from its immediate predecessor which consisted of three separate syllabuses each catering for a different range of ability. Each syllabus commenced with a preamble followed by a list of topics. The syllabus was accompanied by notes to the syllabus which in about 90 pages further expanded the content and suggested the number of periods to be spent on each topic. For most teachers, the 90 page expansion effectively became the Syllabus.

After the present Syllabus had been officially in operation for a year, the Syllabus Committee invited the Centre for Research in Measurement and Evaluation to conduct an external evaluation of the Syllabus. An evaluation subcommittee was appointed from within the Syllabus Committee to act in an advisory capacity only to the Centre which was to have responsibility for designing and implementing the study. This paper deals primarily with two of the methods used in the evaluation; namely, interviews and questionnaires; and with the results obtained using questionnaires. Other sources of information used in the study but not dealt with in this paper were copies of junior science programs and diaries kept by teachers of their classroom activities.

The choice of an evaluation model

As Hamilton (1976) points out, the choice of an appropriate Evaluation Model depends very much on the type of syllabus being evaluated and the questions that can sensibly be asked about a syllabus.

For example, a highly content biased Syllabus whose primary aim was to instil particular items of knowledge could probably best be evaluated by knowledge tests designed to assess the achievement of objectives. To do so would be to adopt the model espoused by Popham and Baker (1970). By contrast, the Syllabus at hand is one which fits Hamilton's category of situationally biased syllabuses. Syllabuses falling in this category aim to produce programs that accurately reflect the life world of the child. They do not rely on detailed specification of content and organisation, but rather upon a flexibility of interpretation and implementation. By so doing they are likely to generate diversified programs with different objectives. All these characteristics imply that an evaluation of the Syllabus on the basis of say state-wide achievement tests would be inappropriate.

Accordingly, the evaluation subcommittee opted for an evaluation which concentrated on the implementation of the Syllabus. After all, in terms of evaluating the Syllabus, it is immaterial what the student outcomes are unless one can first show that the Syllabus is being implemented.

In terms of Stufflebeam's CIPP model of evaluation for decision making (Stufflebeam, 1971) an emphasis on implementation of the Syllabus focuses more directly on context, input and process rather than on product. In terms of Stake's (1967) model of evaluation, the emphasis is on antecedents and transactions rather than on outcomes.

Terms of reference for the study

After several discussions involving members of the evaluation subcommittee, a number of questions emerged which were to steer the course of the study. These questions were:

1. To what extent are teachers aware of and making use of the Statement of Aims and Content prepared by the Junior Science Syllabus Committee?
2. What do teachers perceive to be the strengths, weaknesses and ambiguities in the Syllabus?
3. How are teachers implementing the Syllabus?
4. What types of support do teachers feel they need in order to implement the Syllabus?

By asking these questions the subcommittee hoped to obtain information that would be useful to the Syllabus Committee in:

1. Drafting modifications to the Syllabus to submit to the Secondary Schools Board for approval. Such modifications were envisaged as part of the ongoing development of the Syllabus. There was no intention of replacing the Syllabus with something radically different either in philosophy or in substance.
2. Identifying inservice education needs and making recommendations for changes in inservice education where necessary.

At this point it could not be said that there was any specification of decisions to be made. Rather, it was expected that in the course of the study issues would emerge, problems would be identified and only then would it be possible to say what decisions had to be made.

The evaluation plan

Once the major questions to be addressed by the study and the purposes of the study had been agreed upon a plan of action was drawn up. It would have been possible simply to send a questionnaire to a sample of teachers throughout the state asking their opinions on the Syllabus. However, by so doing the study might easily have bypassed the main issues of concern to teachers or have dwelt on issues which were of no concern to teachers. Hence it was necessary to have teachers identify the issues themselves. In addition, the subcommittee sought information on what was actually happening in the schools. Accordingly, it was agreed that a number of schools would be studied fairly intensively using a variety of methods to gather information. The study of these schools was to provide information for the construction of relevant questionnaires to be sent to state-wide samples and to develop case studies on those schools.

Mainly for practical reasons it was decided that the intensive study of schools would take place in 2 of the 11 regions in N.S.W., one city region and one country region. Within each region 20 high schools (or central schools with a secondary department) were chosen so that the sample included schools ranging from very small central schools with a secondary department and one science teacher to large high schools with perhaps twelve science teachers. It is possible that the implications of the present Syllabus would be very different for each of these two extremes. The aim was to canvass the range of opinions about and approaches to the Syllabus rather than to obtain information on the frequency of occurrence of such opinions and approaches. The sample of schools was not therefore directly proportional to the population. To sample proportionally would probably have been to submerge minority opinions. In terms of Stake's model of responsive evaluation (Stake, 1976) the intention was to identify "the whole constellation of values" held for the syllabus and "the complex of expectations and criteria that different people have for such a program".

The interviews

Each school was visited for one day by either myself or one of two science masters in the region who had been released for a week to help with the study. Before commencing to work in the schools on their own each science master accompanied me to a school for a briefing session and practice run on interviewing. The fact that five people were involved in interviewing helped to ensure against interviewer bias.

An interview schedule dealing with a number of areas was piloted at four Sydney high schools. Since it was desirable to restrict the length of an interview to 35 minutes, or one school period, it was found to be necessary to divide the original interview schedule into two.

One interview schedule dealt primarily with issues of programming and assessment, the impact of various aspects of the Syllabus on teaching, strengths and weaknesses of the Syllabus and possible changes to the Syllabus. This interview schedule, because it dealt with changes since the introduction of the Syllabus, was suited to teachers with three or more years of teaching science in N.S.W. Sixty-five interviews were conducted using this schedule.

The other interview schedule dealt with awareness of and use of the Syllabus, operational aspects of various aims, clarity of the aims, problems in using the Syllabus, the usefulness of inservice activities and possible improvements to inservice, strengths and weaknesses of the Syllabus and possible changes to the Syllabus. One hundred and ten interviews were conducted using this schedule. The interview schedules, although structured, were only seen as a stimulus for obtaining information and for encouraging teachers to discuss their experiences in using the Syllabus.

Many of the issues which arose frequently in the course of interviews appear in the questionnaires and will be discussed in conjunction with the questionnaires. Other issues which emerged in the interviews but were dealt with only briefly if at all in the questionnaires were an alleged increase in the involvement of rank-and-file science teachers in programming; an "acceptance" of the aims by teachers as self-evident often, however, without any real understanding of the aims; a continued tendency by teachers to equate the Syllabus with a content statement and consequently to see the present Syllabus as a licence to teach anything of preference to the teacher. As one science master remarked: "I could teach the thoughts of Chairman Mao and justify it in terms of that Syllabus". There appeared to be some resentment among teachers, and science masters in particular, that they were being asked to perform an additional duty, namely that of constructing a curriculum, with no extra time to do so. One science master commented that he was being asked to be both the architect and the bricklayer and that by necessity he was laying fewer bricks.

Developing the questionnaires

From the interviews a number of recurrent issues became evident. Each of these issues subsumed a number of related concerns. For example one issue involved the brevity of the statement of content in the Syllabus. A related concern was the possibility that students might miss out on the basics of science. The issues which emerged formed the framework for three questionnaires, originally a single questionnaire which proved in a pilot study to be too time consuming. Three further pilot studies were required before the questionnaires were considered to be satisfactory.

Questionnaire 1 dealt with the purposes of junior secondary science as teachers think they should be and as they are implied by the Syllabus. It also included a brief section on the supporting documents to the Syllabus.

Questionnaire 2 dealt with the Statement of Content in the Syllabus and the various arguments for and against its expansion.

Questionnaire 3 dealt with the Attitude Aims in the Syllabus and teachers' reasons for their appraisal of the attitude aims. This questionnaire also included a section on the assessment of attitudes and the various arguments for and against the assessment of attitudes for the purpose of certification.

All three questionnaires (for motivational reasons) included an identical section on support for teachers. The inclusion of this section was to stress that we wanted teachers to help us to help them.

In writing the questionnaires every effort was made to use statements that had been made by teachers during the interviews. By so doing it was less likely that the questionnaires would include wording or ideas that were unfamiliar to teachers or appeared stilted.

The samples for the questionnaires

From the total population of science teachers at government secondary schools three random samples were drawn without replacement, one sample for each of the three questionnaires. There were 425 teachers in each sample. This allowed each sample a reliability of ± 5% (for a 95% confidence interval). The necessary sample size for such reliability is about 350. Hence the sample size of 425 chosen allowed for the inevitable non return of some questionnaires.

Questionnaires were mailed directly to the selected teacher with an addressed return envelope. Questionnaires were coded for follow-up purposes but the confidentiality of answers was stressed. A week after sending the questionnaires a reminder note in the form of a cartoon was sent to those who had not at that stage returned their questionnaires. A week after the first reminder a second more formal reminder letter was sent to those who still had not returned their questionnaires. By the time the questionnaires were to be ready for data punching the following return rates had been attained:

Questionnaire I – 79%
Questionnaire II – 74%
Questionnaire III – 76%

Major results

Questionnaire I

One questionnaire included a section on the purposes of junior secondary science. For each of ten possible purposes teachers were asked to indicate for both above average and below average ability students separately whether that purpose was essential, desirable or should not be actively pursued. This information was to be used to indicate the degree of congruence between teachers' aspirations for their students and the purposes of science education as they are set out in the Syllabus. Teachers were also asked to indicate whether they considered each purpose to be implied by the Syllabus. Their answers to this question were to be used as an index of their awareness of the Syllabus.

The purposes included in the questionnaire were "Communication Skills", "Future Academic Needs", "Laboratory Skills", "Practical Skills Outside the Laboratory", "Prediction and Control", "Scientific Knowledge", "Scientific Methods", "Social Responsibility of Science", "Understanding" and "Vocational". Each purpose was accompanied by a brief explanation.

Overall, the difference in teachers' priorities for the above and below average ability students is congruent with the emphasis in the Syllabus on adapting teaching to the different needs, interests and abilities of students. "Communication Skills" was rated essential most frequently for both above average and below average ability students and "Understanding" was high on the list for both ability levels. "Vocational" was low in both lists. Aside from these similarities, however, there was little further correspondence between the ratings of purposes for the two ability levels. (Calculation of Kendall's tau for tied ranks produced tau = .44). The emphasis for below average ability students tended to be on everyday applications of science and coping with the environment. For above average ability students scientific methods and laboratory skills were seen as being of high priority.

One point of interest was that there appeared to be far less consensus about what are essential purposes for below average ability students than for above average ability students. Further there was a higher correspondence for the ranking of purposes according to the percentage of teachers who rated them essential for above average ability students and the ranking of purposes according to the percentage of teachers who indicated they were implied by the Syllabus (Kendall's tau for tied ranks = .80) than there was for the same ranking for below average ability students and the same ranking of purposes according to whether they were seen as implied by the Syllabus (Kendall's tau = .56). These data suggest that the Syllabus, as teachers see it, caters better, at least in theory, to above average ability students than to below average ability students.

Questionnaire II

When asked whether they considered that the Statement of Content in the Syllabus should be more detailed than at present 64% of teachers replied in the affirmative. Irrespective of whether the content statement were to be expanded or not the majority of teachers (about $^2/_3$) indicated that it should not be prescriptive.

During the interviews a number of arguments both for and against a more detailed statement of content were offered. These arguments were included in the questionnaire as statements about the effects of the present statement of content or alternatively of a more detailed statement of content. There were forty-three such statements. Teachers were asked firstly to agree or disagree with each of these statements and then to indicate whether their agreement or disagreement with the statement was a reason for their opinion on whether the Statement of Content should be expanded. From these responses it was possible to identify the most popular reasons for a more detailed statement of content. Knowledge of these reasons should give further direction to possible courses of action in relation to the Statement of Content.

Those statements (or their negation) which were endorsed as reasons by around 50% or more of teachers who wanted a more detailed statement of content are presented below. A (+) beside a statement indicates that agreement with the statement was a reason for more detail while a (-) indicates that disagreement was a reason. The percentage in parenthesis indicates the percentage of those wanting more detail who indicated that their agreement or disagreement with the statement was a reason for more detail.

1. A more detailed statement of content is necessary for state-wide assessment. (+) (85%)
2. With the present Statement of Content students could miss out on some of the basics of science. (+) (82%)
3. A more detailed statement of content is necessary to allow comparability between schools on the basis of achievement of students. (+) (82%)
4. A more detailed statement of content is necessary if teachers are to comply with aim 2.1 of the Syllabus – "To establish a store of facts, ideas and laws which are important in understanding the nature and behaviour of man and his environment." (+) (79%)
5. Those teachers who prefer a more structured approach should be given the option of relying on a more detailed statement of content. (+) (74%)
6. Teachers need help in selecting content. (+) (73%)
7. Students should be able to transfer from one school to another with minimal disruption to their progress. (+) (69%)
8. A state-wide junior secondary science syllabus should outline only the aims of science education. (-) (69%)
9. The present Statement of Content gives no advice on the needs of students continuing to years 11 and 12. (+) (68%)
10. There should be greater uniformity of content taught throughout the state than at present. (+) (67%)
11. With the present Statement of Content high ability students are not being given sufficient depth of knowledge. (+) (64%)
12. The increased emphasis on attitudes, skills and understanding in the Syllabus makes a more detailed statement of content unnecessary. (-) (60%)
13. The move away from an emphasis on knowledge in the Syllabus has led to a depreciation in the quality of education. (+) (60%)
14. Most teachers know what basic ideas should be covered without having them spelt out. (-) (60%)
15. The present Statement of Content allows a greater amount of effort to be concentrated on some disciplines than on others. (+) (57%)
16. It should be the responsibility of the science department of a school to outline

content rather then the responsibility of the Science Syllabus Committee (–) (56%)

17. The present Statement of Content encourages a teacher to pursue his/her interests in selecting content. (+) (55%)
18. The present Statement of Content has decreased the motivation of students to achieve. (+) (55%)
19. The Statement of Content in the Syllabus is not a sufficient basis on which to build a program. (+) (50%)
20. With the present Statement of Content it is very difficult to reach staff consensus on what should be taught. (+) (50%)
21. The present Statement of Content encourages teachers to rely on the old syllabus for guidance on content. (+) (49%)
22. With the present Statement of Content a school's science program could be biased by local considerations. (+) (48%)
23. Teachers lack confidence in their ability to select the right content for their students. (+) (48%)

A visual scan of the reasons most frequently endorsed by teachers for a more detailed statement of content suggested three main themes.

One theme concentrated on the need for a more detailed statement of content for state-wide comparability in assessment and for greater uniformity. These arguments were endorsed in spite of the move toward internal assessment for certification in N.S.W. (see statements 1, 3 and 10.)

A second theme involved an apparent preoccupation with the possibility that students might miss some of the basics. (See statements, 2, 7, 9, 11, 13, 15, 18, 22.)

A third associated theme deals with the difficulties encountered by teachers in selecting content and teachers' lack of confidence that they are including "the right things" in their programs or lessons. (See statements 4, 5, 6, 14, 16, 19, 20, 21, 23.)

The questionnaire included a number of supposedly positive statements about the Syllabus. Some of these statements received agreement from a high percentage of both those wanting and those not wanting a more detailed statement of content. Examples of such statements are:

1. A school should be encouraged to select the content which is most appropriate to its local community.
2. The exercise of selecting content for his/her lessons contributes to the professional development of the teacher.
3. The present statement of content encourages staff collaboration in developing a program.

However, most of those teachers who wanted a more detailed statement of content and who agreed with the above type of statement indicated that their agreement was irrelevant to their opinion that the Statement of Content should be more detailed. This combination of responses together with the interview data suggests that while many teachers applaud the principles of freedom in curriculum development they do not feel adequately equipped either in time or in expertise to cope with such freedom.

Questionnaire III

One feature of the new Syllabus is its division of aims into three sections: Interests and Attitudes, Knowledge and Understanding, and Skills. Although the previous Syllabus included attitudes and skills aims they were intermingled with the Knowledge aims. The Syllabus Committee felt at the time of writing the new Syllabus that the attitudes and skills aims had been neglected in teaching. Accordingly it decided to give these aims more emphasis by expressing the aims under the three separate headings.

During the interviews the desirability of the attitude aims was called into question. Accordingly the questionnaire included some questions on this issue. 89% of teachers considered that the Syllabus should include some attitude aims and 87% thought that the current attitude aims were at least satisfactory. The support was not however an unqualified one since 60% of teachers felt that the attitude aims should be modified. The interview data gives some ideas as to how they might be modified.

The main attributes of the attitude aims perceived by teachers were that they emphasised many aspects of the student's development, that they emphasised the relationship of science to life, man and society and that the attitude aims allow teachers to cater more to the interests and abilities of students.

Another issue which arose frequently in conjunction with attitudes was the assessment of attitudes for various purposes. On the basis of questionnaire data 43% felt that attitudes should be assessed for the award of the school certificate, 63% said assessment of students should be used diagnostically and 73% said that assessment of attitudes should be used to provide feedback on the effectiveness of teaching. Only 17% said that attitudes should not be assessed for any purpose.

With the increased emphasis on attitudes in the Syllabus the Syllabus Committee has advised schools to include assessments of attitudes in their school certificate assessments for each student. The school certificate grade in science is to be an aggregate of marks for interests and attitudes, knowledge and understanding and skills. The inclusion of attitude assessments in the school certificate proved to be a controversial issue during the interviews. The ambivalence within the teaching population is reflected also in the results of the questionnaire. 43% either agreed or strongly agreed with such assessment, 44% either disagreed or strongly disagreed with such assessment and 13% were undecided. The question appears to have virtually polarised the teaching population.

Owing to the degree of controversy and the fervour with which the issue was raised in the interviews teachers were asked (in the same way as they were for the question of a more detailed statement of content) to indicate their reasons for their particular opinion on whether attitudes should be assessed for certification. The arguments for assessing attitudes most frequently endorsed by teachers were that the assessment of attitudes allows a sense of accomplishment to those students who do not excel in knowledge and/or skills; that students should be rewarded by marks for their accomplishment of attitude aims as they are for knowledge and skills aims; that students should be assessed on all Syllabus aims. The arguments against assessing attitudes most frequently endorsed were that it is at present impossible to assess attitudes with a reasonable degree of reliability and that attitude assessments are considerably influenced by teacher-student relationships.

Support for teachers

As mentioned previously all three questionnaires included a section on support for teachers. In order to obtain a better estimate of the opinions of the population than would be allowed by any one of the three samples on its own, combined scores were obtained using weighted percentages from each sample. The percentages were weighted according to the number of respondents.

The questionnaire includes thirty-six types of support each of which teachers were asked to rate as essential, highly desirable, desirable, unnecessary or undesirable. There appears to be little consensus among teachers as to which types of support are essential. Only four types of support were rated as essential by more than 20% of the population. These were:

1. Concrete advice on how to make science relevent to low ability students. (Rated essential by 27% of respondents)
2. A list of available resources and their possible uses. (Rated essential by 25% of respondents.
3. Advice on the assessment of attitudes. (Rated essential by 22% of respondents)
4. Information on the basics required by students progressing to years 11 and 12. (Rated essential by 21% of respondents)

It is interesting to note that for each of the latter two types of support almost as many respondents rated them as unnecessary or undesirable as rated them essential.

In order to identify those items which were most frequently rated as at least desirable the percentage of respondents rating each type of support as desirable, highly desirable or essential were summed. The ten most popular appeared to revolve around two major themes. The most clearly evident theme related to provision of resources or information on resources. Every item dealing directly with resources in the questionnaire was included in the ten most popular types of support. The items referring to resources were, in order of popularity:

1. A list of available resources and their possible uses. (98%)
2. Access to a central pool of ideas and resources. (97%)
3. Access to a resource file for each content area of the Syllabus. (94%)
4. A list of recommended texts with information on their strengths and weaknesses. ([illegible]%)

(Percentages in parentheses refer to percentage of teachers indicating that that type of support was at least desirable.)

A second theme which emerged in the top ten types of support might loosely be referred to as dealing with inservice needs. Since a large proportion of the 36 items deals with inservice needs it is not surprising that such items appeared in the top ten. The items referring to inservice were in order of popularity:

1. Concrete advice on how to make science relevent to low ability students. (94%)
2. More opportunities for discussion with other teachers about approaches to implementing the Syllabus. (93%)
3. Refresher courses on new equipment. (91%)
4. Advice on possible alternative approaches to implementing the Syllabus. (90%)
5. Help in developing interest topics. (90%)
6. Advice on how to deal with mixed ability groups. (88%)

All forms of teacher support are considered to be at the least desirable by 50% or more of the teachers. This probably reflects an acquiescance to suggestions for help on behalf of teachers rather than genuine need.

Summary and conclusions

This paper has reported on two of the methods used to evaluate the implementation of the N.S.W. Junior Secondary Science Syllabus. The study was designed to address a number of questions which formed the framework for the construction of interview schedules. The data obtained using the interview schedules in turn provided the basis for the construction of questionnaires pertinent to science teachers in N.S.W.

The questionnaire data reinforced by the interview data suggest that teachers are generally favourable to the notion of a syllabus which allows teachers freedom of choice in selection of content *provided* that they are given the support necessary to take advantage of this freedom with confidence.

This study is seen as an integral part of the cycle of Syllabus development. It is distinctly formative rather than summative evaluation. (See Scriven 1967.) Regular and frequent communication with the teaching force is vital to the formation of a Syllabus which is in touch with teachers' needs as curriculum developers and implementers.

References

HAMILTON, D. *Curriculum evaluation.* London: Open Books Publishing Limited, 1976.

POPHAM, W.J. & BAKER, E.L. *Systematic instruction.* Englewood Cliffs, N.J.: Prentice-Hall, Inc., 1970.

SCRIVEN, M. *The methodology of evaluation.* A.E.R.A. Monograph Series on Curriculum Evaluation No. 1. Chicago: Rand McNally, 1967.

SECONDARY SCHOOLS BOARD, New South Wales. *Syllabus in Science. Years 7-10 Statement of Aims and Content.* N.S.W.: Government Printer, 1975.

STAKE, R.E. The countenance of educational evaluation. *Teachers College Record*, 1967, 68, 523-540.

STAKE, R.E. Programme evaluation, particularly responsive evaluation. In Dockrell, W.B. & Hamilton, D. (eds.). *Rethinking educational research.* London: Hodder and Stoughton, 1976.

STUFFLEBEAM, D. (ed.) *Educational evaluation and decision making.* Itasca, Illinois: Peacock, 1971

Research in Science Education, 1977, 7, 61-70.

SCHOOL FACTORS AFFECTING CURRICULUM CHANGE IN SCIENCE

John M. Owen

Introduction

During the last decade, important changes have occurred in the planning of curricula in Australian secondary schools. Teachers in all states now have more responsibility for the courses they offer than they did ten years ago. Devolution of control to schools from central authorities has been manifested through the abolition or reduction in the amount of external assessment, the replacement of prescriptive syllabuses by documents of an advisory nature, and the increased importance of consultancy services to schools coupled with the demise of teacher assessment by inspectors. The changes have, so far, affected the lower levels of the secondary school to a greater extent than the upper levels.

Between 1969 and 1974, the Australian Science Education Project developed materials which were designed predominantly for use in school Science classes in Years 7-10. Forty-eight units were developed, including one intended for use in teacher education. The project was the first national curriculum development project in Australia.

The units were arranged into three levels of difficulty, based on Piaget's theory of intellectual development. The form of the materials was designed to encourage teachers to use student inquiry as the basis of learning, and to provide scope for students of varying abilities to work at their own rates, by studying alternative segments, or options. Each option was also classified according to its reading and intellectual levels. Writers of the units tended to favour a cross-disciplinary approach to Science rather than the traditional subject areas. An important additional feature of the project was that the ASEP materials were written and packaged in unit form, so that they could be used in varying amounts in school courses, depending on the needs of the teachers. Altogether the ASEP materials represented a significant innovation in teaching Science in Australian schools, and it was of interest to discover the extent to which the materials were used in schools, and the problems associated with the introduction of the units into schools. This paper is devoted to a brief examination of the latter of these two aspects, that is the problems associated with the introduction of the materials into schools.

A unit of analysis

A necessary prerequisite to the collection and analysis of information relating to the implementation of ASEP materials in schools was to decide upon a suitable unit of analysis. Taylor (1970), showed that, in the United Kingdom, groups of secondary school teachers were continuously involved in the planning of courses when the subject was taught over a number of year levels.

In Australia, two recent studies were concerned with the planning of Science courses. Beeson and Gunstone (1975), in a survey of Victorian schools reported that of the 52 per cent of all science coordinators who responded to a questionnaire, just over one half said that all staff teaching Science were involved in decisions about the courses. In another instance, Northfield (private communication) found that in two thirds of the schools contacted, the staff as a group were responsible for planning the Science syllabus.

This evidence, coupled with the knowledge of trends in Australia toward school-based planning involving all teachers in a subject area, led to the selection of the group of Science teachers in a school as the unit of analysis in the study.

Literature review

Very little work appears to have been carried out in the field of monitoring curriculum change in schools. The studies which have been reported here employed a range of methods, and have been carried out using various units of analysis, such as individual teachers, groups of teachers, the school and the school district. The major forms of data collection have been surveys or case studies. The paucity of reports of studies in science education led to a survey of innovation in other subject areas of education in order to form a basis for data collection. From the review, five characteristics of the user system which appeared to affect the successful adoption and use of innovations were defined and labelled interaction, organization, cooperation, openness and knowledge. The definitions are presented below.

Studies by Jenkins (1967) and Yegge (1971), Nicodemus and Kelly (1973), Hensel (1974) and Corwin (1975) emphasised the need for linkage between external expertise and teachers engaged in implementing innovations in schools. In a national survey of the factors affecting the use of innovations in schools throughout the U.S.A., Havelock (1973) included two terms 'linkage' and 'proximity' which both related to communication. This element was labelled interaction.

> *Interaction* is the extent to which members of the Science staff in a school are linked to other users, to others with greater expertise in the use of relevent curriculum materials, and those in the educational milieu concerned with the dissemination of ASEP materials.

Gross *et al* (1971), in a study of an innovative program in a single school showed that a failure of management was significant in the failure of the innovation. Overviews at various levels of generality have also considered organizational factors. Williams (1974/5) reported that effective management was important in explaining the success or failure of the Nuffield Science programs in schools. Havelock (1973) included a term pertaining to efficient use of internal resources in his overview of the factors affecting the efficient flow of knowledge to receivers. The nature of the ASEP materials was a *prima facie* case for including such an element in an investigation of their use in schools. This element was named organization.

> *Organization* is the extent to which the resources of the Science department and its facilities in a school are effectively deployed by the teachers of Science in the school.

A third element related to staff morale. Ochitwa (1973) showed that the 'esprit' of high adopting schools was significantly higher than for low adopting schools in a study of innovation in Canadian primary schools. Smith (1971), Benson (1973) and Runkel (1974) also pointed out need for effective internal communication amongst groups considering innovatory programs. The third element was labelled cooperation.

> *Cooperation* is the extent to which staff teaching Science in a school assist each other in the day to day functioning of the department and the degree to which decisions are made which involve staff as a whole.

Fourthly, there was an element which described the users attitude to change. Watson's (1966) overview suggested that conformity and adherence to traditional activities were significant inhibitors to change. Havelock's (1973) term was openness and is defined in terms of reaching out for new ideas and new products. This term was adopted for this study.

> *Openness is the extent to which staff teaching Science in a school express a willingness to find out about new developments, a willingness to be helped, and a desire to change.*

A fifth factor also kept occurring in the literature. Gross *et al* (1971) and Kelly and Monger (1974) showed in studies of educational change that the user's lack of clarity of the innovation was an important factor in its rejection. Northfield (1975b) suggested that, in the case of the dissemination of ASEP materials many teachers had an inadequate *knowledge* of the properties of the materials. These five elements were seen as useful underlying characteristics around which data could be collected from schools.

Data collection and analysis

For each of the elements just defined, a set of statements was written and trialled before being used in the large scale data collection (Owen 1977). These statements were included in a questionnaire which was directed at all Science teachers in a sample of 290 secondary schools in all states of Australia. A 90 per cent school return rate was achieved. A school score for each of the scales; interaction, organization, cooperation and openness and a score of the knowledge of teachers about ASEP was obtained by aggregating the responses of all teachers of Science in each school. Table 1 lists the statistics of the four scales.

TABLE 1

Scale statistics for view scales

	Interaction	Organization	Cooperation	Openness
Between Teachers				
Number of items	7	7	6	6
Mean	12.9	23.1	20.3	21.9
Standard deviation	5.32	4.46	4.27	3.03
α reliability	0.71	0.64	0.70	0.60
Between Schools				
Mean	12.9	23.1	20.3	21.9
Standard deviation	2.50	3.21	3.01	1.46
α reliability	0.64	0.77	0.81	0.62

The teacher in charge of Science completed a second questionnaire which provided information about the extent of purchase of ASEP materials. Data concerning the use of curriculum materials in 1975, including the use of ASEP units was provided by teachers.

An analysis of this data showed that significant differences occurred in the mean use of ASEP units between states (Owen 1976). This result suggested that comparisons between the schools relating to the implementation of ASEP materials would be more meaningful if they were confined to schools in the one state.

Results in State A

The first step in the analysis was to investigate whether schools with similar characteristics existed in the state. This was achieved by using a cluster analysis. Two major variable types were used in a Hierarchical Grouping Analysis (Veldmann, 1967, 311). They were:

1. the four scale scores and the school score on knowledge of ASEP. and,
2. two measures of purchase and use of ASEP. These were obtained from a set of factor analyses of raw data. The first was a measure of the extent of early purchase and use of ASEP units in 1975. The second was a measure of the schools' continued purchase of ASEP materials into 1976.

In cluster analysis, cases (in this study, schools) are allotted to groups on the basis of the similarity of their profiles. As the number of groups is reduced, the total within groups variance increases. At each step in the amalgamation of groups, the new clusters are formed on the basis of minimum increase in within groups variance. A decision must be made as to the number of terminal groups; in general these are selected before a relatively large increase in total within groups variance occurs. For State A, four terminal groups were selected. The characteristics of the schools in each group were first examined by considering the mean values of the variables used for all schools in the group, and the deviation of the group mean from the grand mean of all schools in the sample, as set out in Table 2. An analysis of variance showed significant differences between the groups on five of the seven variables.

The data on the groups were then subjected to a linear discriminant function analysis. This was carried out to analyse further the nature of the groups formed, and the results were in close agreement with the patterns in Table 2. A further use of this analysis was that it enabled the identification of a school in each group which had characteristics most similar to those of the mean of each group.

TABLE 2

Characteristics of the four terminal groups of schools in the cluster analysis
State A (N = 41)

Variable	Group 1 (N = 17)	Group 2 (N = 5)	Group 3 (N = 8)	Group 4 (N = 11)	sig
Interaction	+	++++	-	-	**
Organization	+	-	-	-	*
Cooperation	-	++	- -	-	**
Openness	+	+	- -	-	ns
Knowledge	- -	++	0	++	**
Early use	- - -	++	- -	+++	**
Continued purchase	-	++	-	+	ns

Note 1. ** $p < .01$ * $.01 \leqslant p \leqslant .05$

2. 0 : group mean within ±0.05 s.d. of grand mean
± : group mean between ±0.05 and ±0.50 s.d. of grand mean
±± : group mean between ±0.50 and ±1.00 s.d. of grand mean
±±± : group mean between ±1.00 and ±1.50 s.d. of grand mean
±±±± : group mean greater than ±1.50 s.d. of grand mean

Visits to schools

In 1976, a visit was made to each of these four schools which were most typical of the schools in each of the four clusters outlined in Table 2. The major purpose of these visits was to examine the transformation of the ASEP materials into the planned curriculum in Science in each school. The use of large scale data collection was seen as a means of reducing the arbitrariness of the selection of schools for more intensive study. The clustering technique showed that groups of schools with similar characteristics existed, and in addition, the profile enabled knowledge of the "innovation climate" and the extent of use of ASEP over time to be known for each school before the intensive study of the planning of their courses was investigated.

The methodology employed during these visits involved the use of a structured interview with the Head of the Science department, using a tape recorder. Other Science teachers, the laboratory assistants and in some cases other teachers were also interviewed. The procedures used in the school visit phase were not as open-ended as those advocated by MacDonald (1971) and Hamilton (1973), although in most schools there were unique factors which contributed to the understanding of the use of ASEP in the school Science programme. The procedures used on collecting data in the schools contained some of the characteristics of those used by Goddlad *et al* (1974). These researchers collected information in a large number of schools after a list of common factors which seemed to be important in explaining differences between schools had been assembled, using the results of other relevant studies.

Similarly, the overall plan of this study attempted to take into account the results of other studies in the field of interest, but at the same time, an attempt was made to build in sufficient flexibility to cater for unique situations in some schools. One writer has eloquently captured the rationale employed here. After reviewing trends in evaluation techniques, he suggested that a proper means of conducting "illuminative' research should

> . . . . begin with the necessary openness and receptivity, holding constricting assumptions in check, but having already assembled on the sidelines resources, in terms of available theoretical orientations which might prove useful in 'making sense' of the situation being investigated (Parsons, 1976, 132).

A brief report on three schools

(The school selected for study in Group One had not used ASEP, and is not included here.)

Group 4: Table 2 shows that the school (which I shall call Cross High School) typical of this group had purchased large amounts of ASEP materials, but was significantly below the state mean on each of the four scales describing the "innovation climate". It was not surprising therefore, to find that the Cross High School's Science department had experienced great difficulty in planning a course in Science in which the ASEP materials were to be the major textual materials.

The choice of ASEP materials for use in the school had been made in 1973, when a new Head of Science was appointed. One of her first decisions was to discontinue the use of the science textbook which was found to have a reading level too high for the vast majority of students. In 1974, a book hire scheme was set up to replace the ownership of individual textbooks. By December 1974 sufficient ASEP materials had been purchased to enable 13 units to be used in the Science course in 1975.

Implementation, however, ran into major difficulties in that year. There was a large turnover of teaching staff; four of the six teachers were new to the school. The chances that these teachers were familiar with new concepts in Science teaching were reduced because three had not undertaken any form of preservice training. In effect, most teachers had to become familiar with the ASEP materials as they were being used, which led to high demands on the time of those involved.

The problems in the school due to the deficient backgrounds of the majority of staff were magnified by organisational arrangements within the Science department. The ASEP materials were not stored efficiently so that they could be quickly assembled for use by teachers. In addition, they were located at the rear of one of the laboratories, which caused problems for teachers when materials had to be moved at the beginning and end of lessons. Also, there was a reluctance to use the laboratories by teachers, even though they were generally available for Science.

The major source of the difficulties, however, lay in the lack of cooperation between the staff involved. The Head of Science was not able to devote all her time to the administration of the department, because of other duties in the school. No courses of study were produced, and teachers had to rely on informal meetings to decide what was to be taught in the coming weeks. There was some resentment among the staff that the Head of Science was not able to provide the support that some teachers needed. A further difficulty related to the relations between the laboratory assistants and the science teachers, because the assistant was either unable or unwilling to assemble the ASEP units for use, and in addition, attempted to impart to the Science teachers his knowledge of traditional science to the teachers which was not asked for and often irrelevant to the situation at hand. In Term I of 1975, teachers largely by-passed the assistant and attempted to organise the ASEP apparatus and textual materials during the first term vacation. The Head of Science also invited a visiting consultant to the school to help in planning a program of study based on the use of ASEP materials. However, the combination of factors of lack of experience, the demands of day to day tasks, and in particular lack of structure of the program and insufficient cohesion between personnel caused the cessation of large scale use of ASEP by the end of 1975 and as a result, a further order for more ASEP materials for Cross High School at the end of 1975 was cancelled.

Group 3: Table 2 indicates that the school (Sherry High School) typical of this group had made no use of ASEP to 1976, and that the school was significantly below the state mean on each of the four scales, and in particular on the scales of 'openness' and 'cooperation'.

An important factor which helped explain this profile was in the transition of responsibility for the administration of the Science department in the period from 1974 to 1975. Up until 1974, most of the Science teachers had been working at the school for at least five years, and for the two senior teachers of the group, Sherry High School had been their only teaching appointment. Courses in Science were predominantly didactic in nature, and although there were three laboratories in the school, an examination of records showed that they were used for less than 10 percent of all class time in Science in 1974. Overall, the department's leadership had been content to continue for a number of years without reviewing the subject matter taught or the methods used in the courses offered.

In 1975, a new Head of Science was appointed to the school, and immediately determined that the structure of the programme should undergo some modifications. He had been strongly influenced by his experiences as a young teacher in a school well known in the state as being one of the first to experiment with individual progression methods. He was also aware of recent innovations in curriculum development and saw the use of textual materials which encouraged the use of student enquiry as more suitable for the students to use than the textbook which had been standard issue to students at Sherry High School for many years.

The Head of Science discovered that there were two major sub-groups of teachers who would need to have guidelines put before them before any major course changes could be made. In addition to the conservative long serving group of Science teachers, a second group of novice teachers was also involved in teaching Science as an 'extra' to their duties in other subject areas.

As a means of stimulating interest the Head of Science prepared a tentative sequential course outline which was based on a laboratory centred approach. This involved a simple booking system for laboratories which was administered by an efficient laboratory assistant. Other teachers involved were requested to suggest changes to the outline but over a period of two months little feedback was obtained by the department head. The lack of cooperation from the experienced group of teachers was related to their disinterest in making efforts to understand and to try the new teaching methods with the majority of students. The inexperienced group of teachers seemed to be content to be guided by the decisions of the Head of Science.

In the absence of any significant dialogue about the proposed changes the Head of Science made unilateral decisions about the structure of courses and the choices of text books and resources. By late 1975, (when the data which was used to produce the profiles in Table 2 was collected) he had assembled notes and guidelines for each year level, and ordered from publishers some resources from which enquiry-based learning could be carried out. However, after consideration of the general attitudes in the department to change, he deliberately introduced only a few units of ASEP materials into the programmes. A remark he made during and interview sums up his reasons for the small amount of these materials purchased. He said:

> I looked at ASEP, especially the inspection packs, then asked questions to other teachers in other schools. I took JSSP . . . I think JSSP is more restrictive than ASEP. I think inexperienced teachers will handle it a lot better at first, but . . . we will eventually get ASEP.

Group 2: In contrast to the school described above, the majority of teachers at Major High School had been heavily involved in activities related to their professional development as Science teachers in the period from 1973 to 1975, a period of high activity in science education in state A.

When two senior teachers assumed responsibility for the department at the beginning of 1974, there was a nucleus of teachers who were, by and large, aware of the possibilities for the development of a suitable course of study for the students who attended the school.

During 1974, an outline course for Science which covered all form levels up to Year 10 was prepared. 'Home-made' individual progressions resource materials had been used in the past as the basis for teaching the subject but there was concern that students were becoming bored with the use of similar teaching strategies for the entire year, and teachers considered that these materials were not of comparable quality to some resources which had recently been produced commercially.

The new course was based on the use of ASEP materials, and some use was made of other curriculum packages. In addition class sets of textbooks were purchased using money from a levy so that students had a resource book for private study.

The decision to make large-scale use of ASEP materials, suggested initially by one of the Heads of Science, was made easier by the general knowledge that teachers had of the materials. The organization of the materials into an overall plan was mapped out, initially by the two Heads of Science, and then submitted to the entire group of teachers for approval. Minor modifications were then made to this plan after a series of meetings which involved all teachers of Science. Finally, staff working at each year level worked on details particular to each level.

Decisions made about the use of materials and the strategies used were often a compromise between the ideas of individual teachers. This appeared to be due largely to the formation of an initial structure proposed by the two leaders, which gave less experienced teachers a framework from which to work. Also, one of the Heads of Science had an excellent working knowledge of most of the curriculum packages and textbooks, and had been involved in the trials of the ASEP materials. As the program evolved in 1975, the staff agreed to the purchase of more ASEP materials, and by 1976, the overall plan of the Science courses was based on laboratory work. This caused some difficulties in organization, which are reflected in the low profile score on organization in Table 2. The major difficulty encountered related to timetabling of rooms and the transfer of apparatus from laboratories in four different buildings in the school's grounds. By 1976, however, these problems had been largely solved, mainly due to the willingness of the teacher of senior physics to take some of these classes in conventional classrooms, making way for Science classes to make more use of the available laboratory facilities.

Concluding comments

Over the past seven years, the Australian Science Education Project has become well known, by name at least, to many Science teachers in Australian schools. However, there is some evidence to show that the majority of these teachers have not had the chance to become completely familiar with the characteristics of the materials (Northfield, 1975).

The three short descriptions presented in this paper showed some of the problems faced in implementing the materials into Science programs. The results suggest that for the ASEP materials to be used successfully in a Science programme, there is a need for the Science staff to be willing to try new curriculum materials, to be 'professionally' oriented in their attitude to teaching and to be willing to work together in the planning of courses based on the use of the units. In addition, there seems to be a need for the most efficient use of available facilities to be made. It appears that the general abilities related to leadership of the Head of Science were important in shaping the attitudes of other Science teachers and organizing situations where teachers would be involved in planning courses.

An implication of this research is that, although some Heads of Science appear to have a natural capacity to fulfil the role of linker and coordinator, others might benefit from refresher courses in general administration, particularly in the areas of course planning and resource retrieval. There is obviously a role for teachers' colleges in the future to offer such courses to senior teachers, as more and more responsibility is devolved to those at the school level.

Acknowledgment

Work on the preparation of this paper was supported by a grant from the Curriculum Development Centre.

References

BEESON, G.W. & GUNSTONE, R.F. The teachers' role in curriculum decisions. *Australian Science Teachers Journal*, 1975, **21**, (*1*), 5-19.

BENSON, J.A. Comparative study of curriculum innovation in practice. Unpublished M.Ed. dissertation, University of Leicester, 1973. Reported in Harding, J.M., Kelly, P.J. and Nicodemus, R.B. The study of curriculum change. *Studies in Science Education*, 1976, **3**, 1-30.

CORWIN, R.D. Innovation in organisations: the case of schools. *Sociology of Education*, 1975, **48**, 1-37.

GOODLAD, J.I. & KLEIN, M.F. *Looking behind the classroom door.* Worthington, Ohio: Charles A. Jones Publishing Company, 1974.

GROSS, N., GIACQUINTA, J.B. & BERNSTEIN, M. *Implementing organisational innovations.* New York, USA: Basic Books, 1971.

HAMILTON, D. At classroom level. Unpublished Ph.D. thesis. Edinburgh University, 1973.

HAVELOCK, R.L. et al. *Educational innovation in the United States. Volume 1: The national survey, the substance and the process.* Ann Arbor, Michigan: Centre for Research on Utilization of Scientific Knowledge, 1973.

HENSEL, J.W. et al. *Constructs for studying change processes.* Microfiche ED 091 815, 1974.

JENKINS, E.W. The attitude of teachers to the introduction of Nuffield Chemistry. *School Science Review*, 1967, **49**, 231-241.

KELLY, P.J. & NICODEMUS, R.B. Early stages in the diffusion of the Nuffield A-level Biological Science Project. *Journal of Biological Education*, 1973, 7, (*6*), 15-22.

KELLY, P.J. & MONGER, G. An evaluation of the Nuffield O-level Biology Course materials and their use. *School Science Review*, 1974, **55**, 705-715.

MACDONALD, B.J. The evaluation of the Humanities Curriculum Project: a holistic approach. *Theory into Practice*, 1971, **10**, 163-167.

NORTHFIELD, J.R. ASEP . . . . Are teachers prepared? *The Australian Science Teachers Journal*, 1975, **21**, (*1*), 21-26.

OCHITWA, U.P. Organisational climate and adoption of educational innovations. *Journal of Research and Development*, 1973, **4**, (*1*), 38-44.

OWEN, J.M. Three years on: the impact of ASEP materials on Australian schools. *Australian Science Teachers Journal*, 1976, **21**, (*3*), 15-24.

OWEN, J.M. *The impact of the Australian Science Education Project on schools.* 1977, (in press).

PARSONS, C. The new evaluation: A cautionary note. *Journal of Curriculum Studies*, 1976, **8**, (*2*), 125-138.

RUNKEL, P. *Conditions for success and failure of organisational development in schools.* ERIC Microfiche ED 088 242, 1974.

SMITH, M.P. Curriculum change at the local level. *Journal of Curriculum Studies*, 1971, **3**, 158-162.

TAYLOR, P.H. *How teachers plan their courses.* London, UK: NFER Publishing Co., 1970.

VELDMANN, D.J. *Fortran programming for the behavioural sciences.* New York, USA: Holt, Rinehart and Winston, 1967.

WATSON, G. Resistance to change. In Bennis, W.G., Benne, K.D. and Chin, R. (Eds.), *The planning of change.* New York: Holt, Rinehart and Winston, 1966, 488-498.

WILLIAMS, V. The management of innovation: a consideration of implementation. *Journal of Applied Educational Studies*, 1974/5, 3 (*2*), 33-35.

YEGGE, J.F. The adoption of an innovation in physics teaching. Unpublished Ph.D. thesis, Harvard University. ERIC Microfiche ED 052 961, 1971.

Research in Science Education, 1977, 7, 71-81.

THE SYSTEMS APPROACH AND ITS APPLICATION TO CURRICULUM RESEARCH

Roger J. Osborne and Peter S. Freyberg

Introduction

Systems science is a developing field of knowledge, which has its origins in mathematics and engineering. Current literature, such as *IEEE Transactions on Systems, Man and Cybernetics* and *General Systems*, shows that the systems approach can provide a common viewpoint for specialists working in different subject areas. Graduate textbooks on systems theory are appearing (for example, Padulo and Arbib, 1974; Bossel, 1976). It is helpful to realize that there are two distinct aspects to the systems approach:

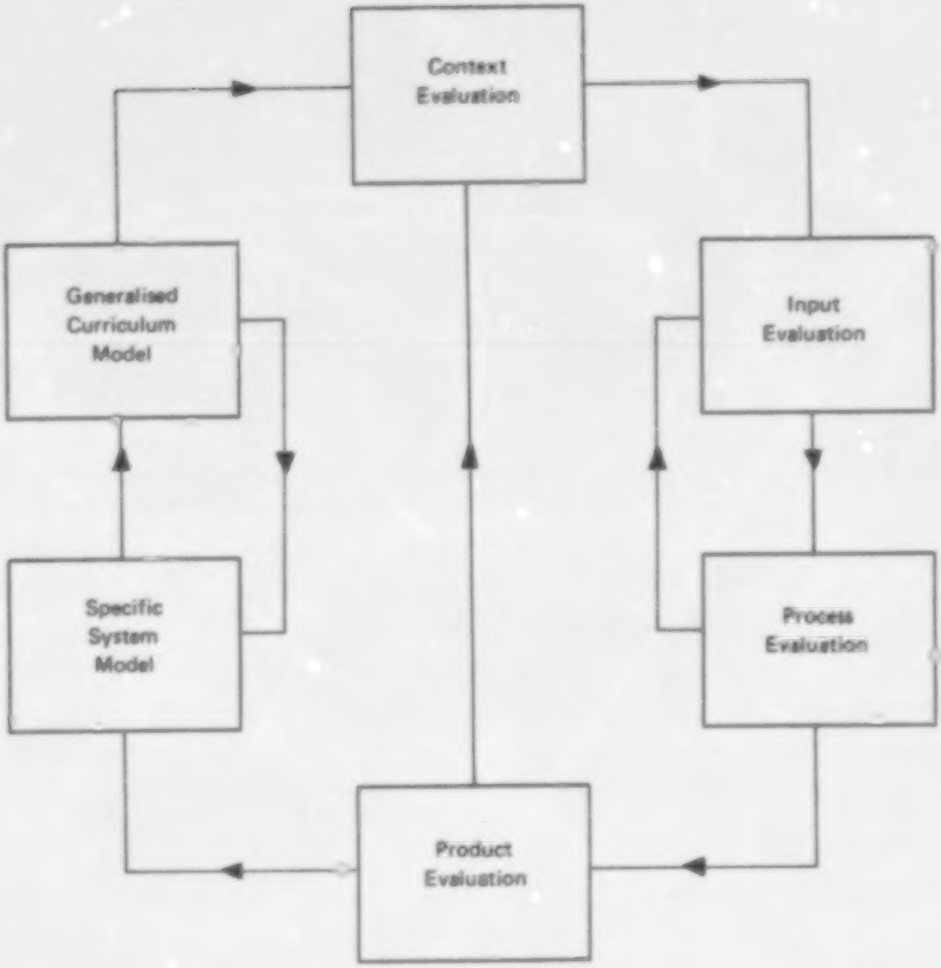
(i) the development of mathematical techniques and methods so that existing systems may be understood in terms of theories and models,

(ii) the development of a systems viewpoint and methodology whereby man-made systems may be effectively and logically developed and refined.

These two aspects of the systems approach will be called the *basic* and *applied* aspects, respectively.

In the field of curriculum development, interest in the systems approach has been almost exclusively in terms of the applied systems approach (see, for example, Lerner, 1973; Rath and Rath, 1973). Unfortunately this has often lead to the defining of objectives and instructional methods with a specificity inappropriate to a non-physical-science and to a non-deterministic system (Yee, 1972). The potential of the systems approach in education has been subject to some rather myopic criticism as a result. As Stowe (1972) has implied, too little emphasis has been given to the basic systems approach. Attention needs to be given to finding out how a system under development and refinement actually operates. As Marchal (1975) has pointed out in his attempt to define the concept of a system, it is necessary to identify system elements and the relationships between these elements.

In terms of the applied systems approach Stufflebeam et al (1971) have developed the CIPP model for curriculum design and evaluation. In Fig. 1 the four basic elements of the CIPP model have been included along with elements based on the basic systems approach, into a broad systems view of curriculum research. It is proposed that rather than concentrate solely on a rigorous applied systems approach more emphasis should be given to isolating, measuring and understanding the relationships among the broader goals and the major procedures of a learning programme. This emphasis implicitly assumes a particular view of the instructional system (which will be called a generalized curriculum model) in which broad goals are of major importance. Once goals, or objectives, have been determined, the applied systems approach can be used to design, develop, and modify the instructional programme. This requires specific development procedures and feedback. Once the instruction has been developed and refined, a specific statistical or mathematical model of the system can be developed which will elucidate the interactions between the students and the instructional system. Hopefully this elucidation will lead to further pertinent and profitable evaluative investigations. At an even more general level the problems of model building should be faced if a real understanding of learning systems, in terms of the purposes for which they have been designed, is to be obtained in the future.


Context Evaluation
Generalised Curriculum Model
Input Evaluation
Specific System Model
Process Evaluation
Product Evaluation

Curriculum Elucidation
(Model Building)

Basic Systems Approach

Curriculum Evaluation
(Development)

Applied Systems
Approach

FIGURE 1

Evaluative and elucidatory components of curriculum development and analysis

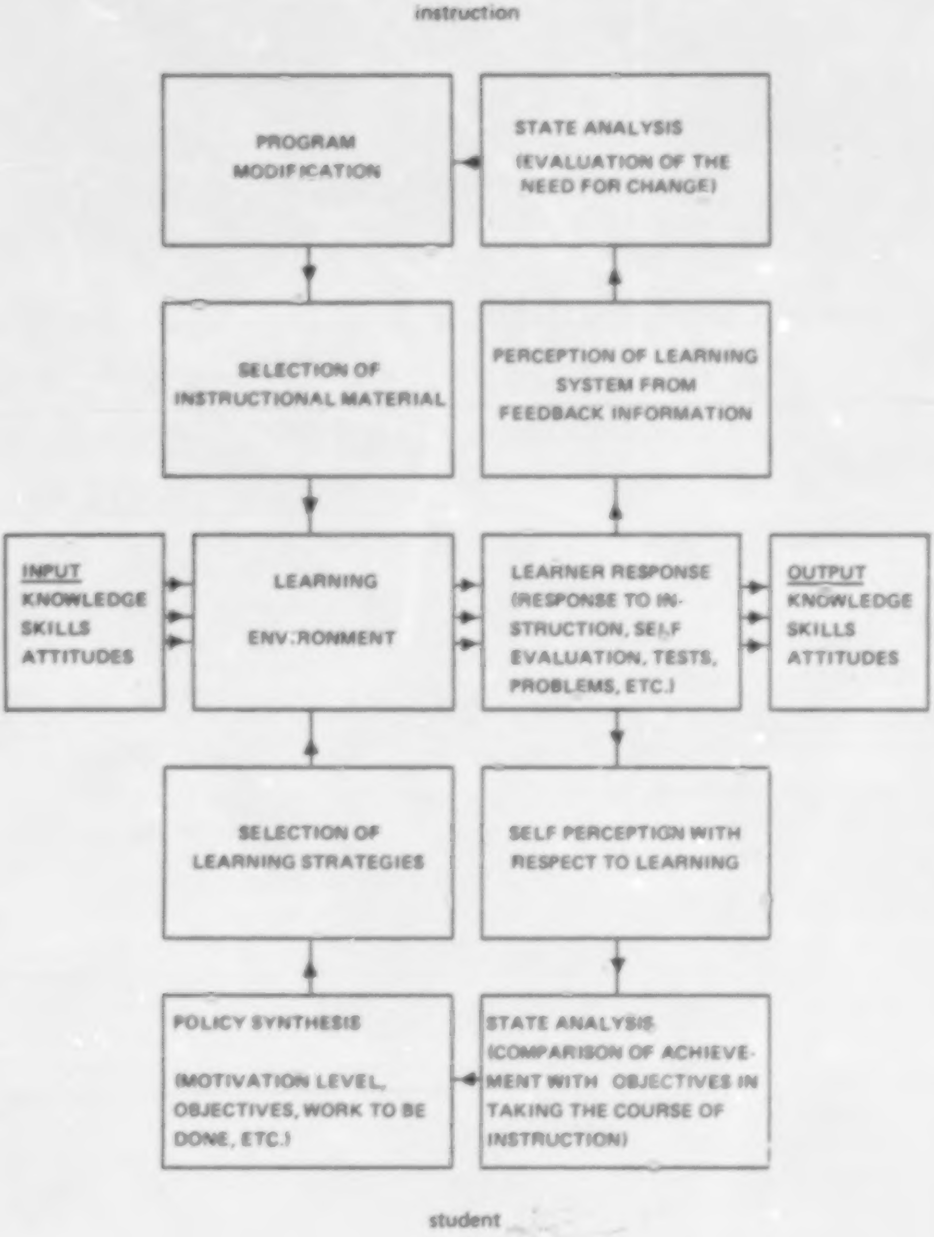

instruction
PROGRAM MODIFICATION
STATE ANALYSIS (EVALUATION OF THE NEED FOR CHANGE)
SELECTION OF INSTRUCTIONAL MATERIAL
PERCEPTION OF LEARNING SYSTEM FROM FEEDBACK INFORMATION
INPUT
KNOWLEDGE
SKILLS
ATTITUDES
LEARNING ENVIRONMENT
LEARNER RESPONSE (RESPONSE TO INSTRUCTION, SELF EVALUATION, TESTS, PROBLEMS, ETC.)
OUTPUT
KNOWLEDGE
SKILLS
ATTITUDES
SELECTION OF LEARNING STRATEGIES
SELF PERCEPTION WITH RESPECT TO LEARNING
POLICY SYNTHESIS
(MOTIVATION LEVEL, OBJECTIVES, WORK TO BE DONE, ETC.)
STATE ANALYSIS (COMPARISON OF ACHIEVEMENT WITH OBJECTIVES IN TAKING THE COURSE OF INSTRUCTION)
student

FIGURE 2

Input-output view of the learning system with emphasis on learner response

A specific application of the systems approach

To this end a conscious effort has already been made to apply, in the development of the first year physics course at the University of Waikato, a truly broad systems approach to course development and analysis. The establishment of this completely new course in a new institution required the clarification of the desired state of the learning system. This was initially clarified in terms of major structural differences between the proposed course and similar courses at comparable institutions. These differences were indicative of a novel emphasis on the importance of laboratory based studies and an increased emphasis on basic understanding rather than rote learning (Liley et al, 1976). Implicit objectives were gradually enunciated in an explicit form. These were subsequently refined and rationalized in the process of teaching and assessing students over a period of two years. The general learner objectives (Table 1) were subsequently found to be most useful for orienting and diversifying both teaching and assessment. These objectives then become central to an input-output systems view of instruction (Fig. 2). This model, combining both teacher and student systems, views learner response primarily in terms of course objectives.

In the first few years of course development, emphasis was placed on the applied systems approach. This involved obtaining feedback to improve teaching, learning and student assessment, particularly in terms of course objectives. In turn, this clarified more clearly the course objectives both to staff and students. Specific questionnaires along with informal student comment, provided feedback information for course improvement. While the absolute values in which students' opinions can be quantified are not particularly useful, comparative values,

(i) for different topics, different experiments and different objectives in the same year (e.g. Fig. 3), and

(ii) for the same aspects in different years (e.g. Fig. 3 and 4), provided suitable data for state analysis (see also Osborne, 1976a).

Once the course had been developed and refined over a period of years, emphasis shifted to the basic systems approach. Attempts were made (Osborne, 1976b) to elucidate the instructional system by developing specific statistical models based on the descriptive generalized curriculum model of Fig. 2.

Among other investigations a major effort was made to investigate student attainment of the course objectives. Input and output student scores on each of the course objectives were obtained using pretested items in test and examination papers supplemented by Likert scale questionnaires. These student scores were analysed using principal components analyses and yielded relatively stable results from one year to the next. Consistently, in analysing both input and output results, four major groupings of the objectives were teased from the statistical information. A typical analysis is shown in Table 2. These investigations showed how the students' attainments with respect to the objectives grouped together for this particular instructional system, for these particular students, and for the particular assessment procedures used. The major groupings were:

THEORETICAL	knowledge and skills to comprehend and analyse physics theory (primarily objectives I and III)
APPLIED	ability to solve algebraic and numerical problems (primarily objectives II(A), II(B), IV(E) and VII(A))

TABLE 1

The first year physics objectives at the University of Waikato

Knowledge and Comprehension of Physics

I	(A)	Develop the ability to extract information (from scientific literature).
	(B)	Develop the ability to interpret information, particularly to interpret mathematical and graphical descriptions.
	(C)	Obtain an understanding of the facts, principles and theories of physics (as specified in the courses).

Application

II	(A)	Develop the ability to apply theoretical results to relatively familiar situations.
	(B)	Develop the ability to apply theory, and mathematically analyze, problem situations.
	(C)	Develop the ability to apply theory to relatively unfamiliar problem situations.

Analysis

III	(A)	Develop the ability to critically analyze problem situations and evaluate information.
	(B)	Develop a critical attitude towards, and an ability to analyze, idealizations, approximations, and assumptions in physical theories.
	(C)	Develop an understanding, and appreciation, of some of the major themes and unifying concepts of physics (including the relationship between fundamental principles and derived laws).

Processes of Experimental Inquiry I (Experimental Design)

IV	(A)	Develop the ability to propose tentative ideas, and working hypotheses, and make logical and testable predictions.
	(B)	Develop the ability to translate problems into experimental operations.
	(C)	Develop the ability to make useful and reliable observations.

Processes of Experimental Inquiry II (Analysis and Conclusions)

IV	(D)	Develop the ability to keep adequately detailed and organised records of experimental work and to be able to extract information from these.
	(E)	Develop the ability to analyze experimental data mathematically and graphically.
	(F)	Develop the ability to draw valid conclusions from observations and data.

Orientation

V	(A)	Develop an awareness of the function and limitation of physical laws.
	(B)	Develop an awareness of the relationship between experimental observation and theory.
	(C)	Develop an awareness of physics as a creative, open-ended and continuing development, both tentative and unfinished.

Attitude

VI	(A)	Develop the attitude that mathematics is a useful, convenient and precise language, and that the use of mathematics in physics is a satisfying activity.
	(B)	Develop the attitude that physics is an important, relevant and satisfying activity.
	(C)	Develop a commitment to actively searching for an understanding of physical phenomena, and for the theories by which they may be explained.

Additional Objective – Comprehension of Mathematical Physics

VII	(A)	Develop the ability to manipulate, and use, the mathematical language of physics.

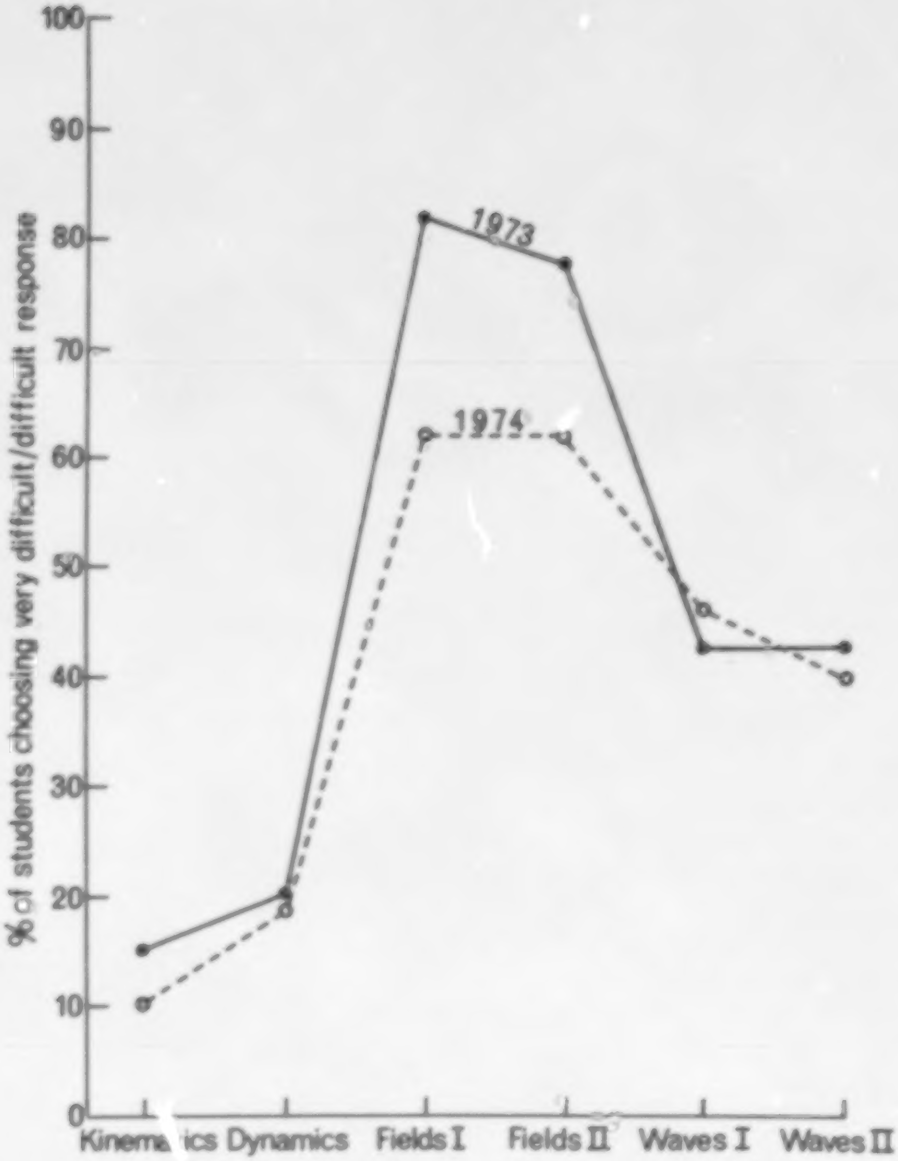


FIGURE 3

Student percentage response on the difficulty on average of the content of sections of the theoretical course. (Very difficulty/difficult versus average/easy/very easy) (N = 80 - 90)

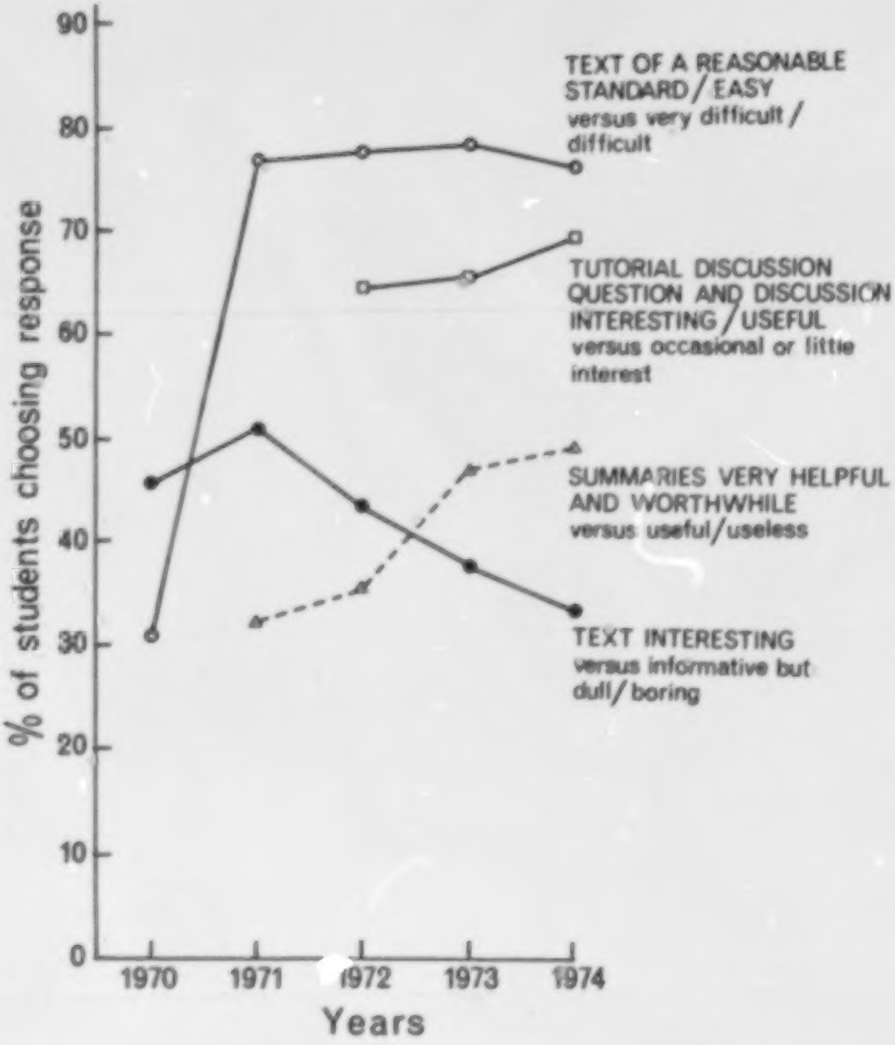


FIGURE 4

Monitoring student attitudes

EXPERIMENTAL knowledge and skills with experimental work (primarily objectives IV)
ATTITUDINAL attitude to physics (primarily objective VI)

TABLE 2

Principal components analyses of objectives scores with factor loadings ⩾ 0.5 shown.
(Varimax rotation, N ≃ 100 for each analysis)

Objective		Question Type	Factor							
			I		II		III		IV	
			1973	1974	1973	1974	1973	1974	1973	1974
THEORETICAL	I (B)	M.C.	0.5	0.5	0.5					
	I (C)	M.C.	0.5	0.7						
	II (C)	M.C.	0.6	0.7						
	III (A)	M.C.	(0.3)	0.7			0.5			
	III (B)	M.C.	0.6	0.7						
	III (C)	M.C.	0.7	(0.3)				0.5		
APPLIED	I (A)	M.C.			0.5	0.7				
	II (A)	M.C.			0.6	0.6				
	II (B)	M.C.		0.6	0.5	0.5	0.6			
	IV (E)	M.C.	0.5		0.5	0.5				
	VI (A)	Likert			0.7	0.7				
	VII (A)	M.C.			0.9	0.6				
EXPERIMENTAL	IV (A)	S.Q.					0.7	0.6		
	IV (B)	S.Q.					0.5	0.7		
	IV (C)	S.Q.					0.7	0.7		
	IV (D)	S.Q.					0.6	0.6		
	IV (F)	S.Q.					0.7	0.7		
	V (A/B)	S.Q./M.C.					0.6	(0.3)		
ATTITUDINAL	V (C)	Likert							0.7	0.6
	VI (B)	Likert							0.9	0.8
	VI (C)	Likert							0.9	0.9

Path analysis (Wright, 1960; Gimmel, 1974) has application in a wide variety of fields and statistical models of the instructional system were developed using this statistical technique. Typical examples of these models are shown in Figs. 5-7. In Fig. 5 attainment of students at the end of the course on the four aspects of achievement identified in Table 2 are related to earlier school attainment in mathematics and physics. In Fig. 6 the four aspects of achievement identified in Table 2 are related to a student's total grade for the entire course (this grade includes 33% internal assessment). In Fig. 7 a student's aim in terms of grades is related to his previous and future attainment in physics.

In the development of models of this type, certain problems were identified. These included

(i) the identification and measurement of important variables,
(ii) the restrictions imposed by any particular mathematical model on the type of

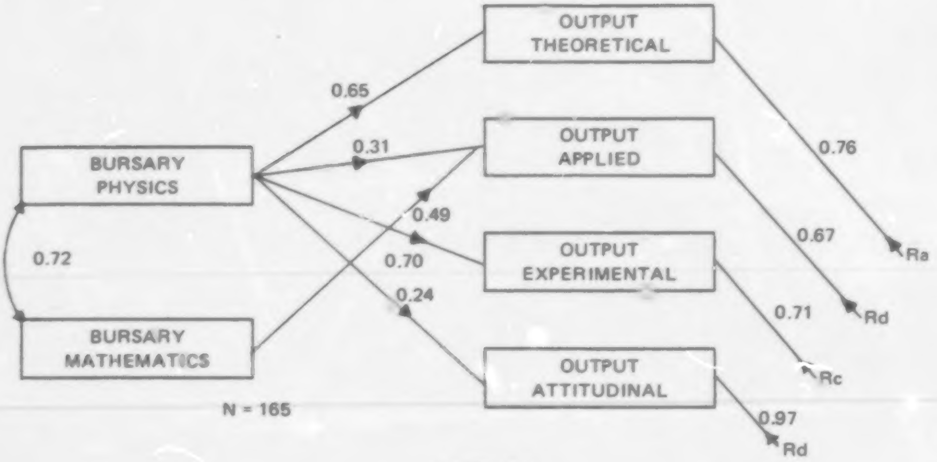

OUTPUT THEORETICAL
OUTPUT APPLIED
OUTPUT EXPERIMENTAL
OUTPUT ATTITUDINAL
BURSARY PHYSICS
BURSARY MATHEMATICS
0.65
0.31
0.49
0.70
0.24
0.72
0.76
0.67
0.71
0.97
Ra
Rd
Rc
Rd
N = 165

FIGURE 5

Path analysis of the first year physics course as an input-output system. (Only significant paths at the 5% level are shown.)

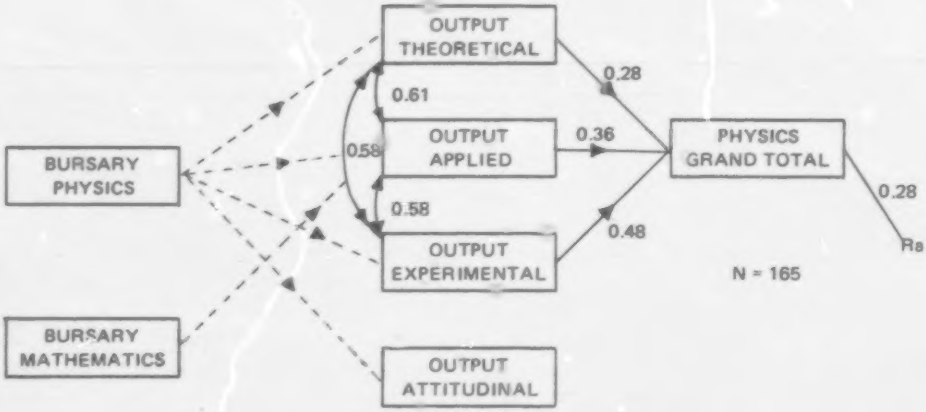

OUTPUT THEORETICAL
OUTPUT APPLIED
OUTPUT EXPERIMENTAL
OUTPUT ATTITUDINAL
BURSARY PHYSICS
BURSARY MATHEMATICS
PHYSICS GRAND TOTAL
0.61
0.58
0.58
0.28
0.36
0.48
0.28
Ra
N = 165

FIGURE 6

The relationship between output scores for particular aspects of achievement and the total grade mark for first year physics (including 33% internal assessment)

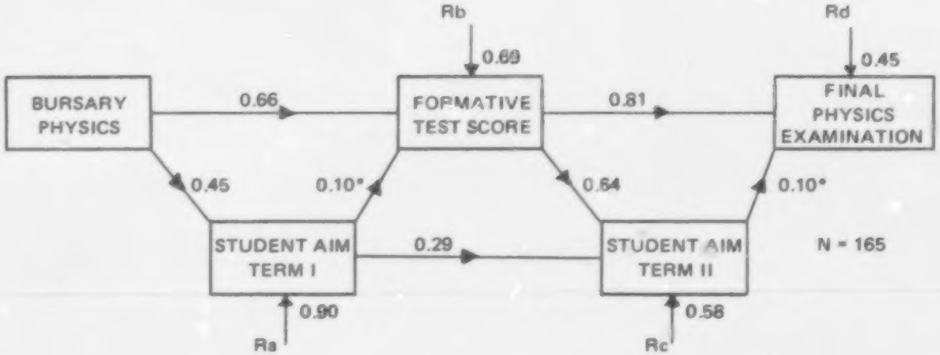


FIGURE 7

Relationships between achievement scores and student aim

data which can be used and the relationships which can be explored, and

(iii) the correct interpretation of the relationships enunciated by the models.

Nevertheless we believe our attempt to develop models, and the models themselves, have given us a clearer appreciation of our instructional and assessment procedures, and thay have led to further specific evaluative studies designed for course improvement.

Conclusion

The application of the systems approach to the development and analysis of a first year university physics course has clarified some of the strengths of this approach as well as some of the limitations and problems. In this particular study the broad systems approach:

(i) provided a comprehensive framework on which to base curriculum work;

(ii) provided a viewpoint which led to an emphasis on particularly important aspects of the instructional system; namely the objectives, the teaching methods, the assessment procedures and the feedback mechanisms;

(iii) ensured that the objectives of the course were clarified and explicitly stated and that possible teaching methods were considered in terms of their potential to achieve these objectives;

(iv) emphasised the importance of obtaining useful feedback information from students so that appropriate changes could be made to instruction and so that the effects of these changes could be monitored;

(v) emphasised the importance of adequate measurement of the attainment of the course objectives both at the beginning, during and at the end of the course;

(vi) emphasised the need for a better understanding of the system through the development of specific mathematical models and provided guidelines for the development of such models; and

(vii) lead to the development of models which suggested further evaluative studies.

In summary, it is considered that this approach can give teachers a more adequate and realistic view of an instructional system and consequently lead to better instruction.

References

BOSSEL, H. et al. (Eds.) *Systems theory in the social sciences.* Basel: Birkhauser, 1976.

GIMMEL, D.R. A path model for course achievement. *The Journal of Experimental Education*, 1974, **43** (*2*), 1-7.

LERNER, J.W. Systems analysis and special education. *The Journal of Special Education*, 1973, **7** (*1*), 15-26.

LILEY, B.S., OSBORNE, R.J. & PEPPER, A.R. A 'laboratory only' course with open book examination. *Physics Education*, 1976, **11** (*6*), 401-404.

MARCHAL, J.H. On the concept of a system. *Philosophy of Science*, 1975, **42**, 448-468.

OSBORNE, R.J. Using student attitudes to modify instruction in physics. *Journal of Research in Science Teaching*, 1976a, **13** (*6*), 525-531.

OSBORNE, R.J. A systems approach to physics education at first year University level. Unpublished D.Phil. thesis, University of Waikato, 1976b.

PADULO, L. & ARBIB, M.A. *Systems theory*. Philadelphia: W.B. Saunders, 1974.

RATH, C., & RATH, G.J., Evaluation is a process; systems analysis in education. *Educational Technology*, 1973, **13**, 60-62.

STOWE, R.A. Research and the systems approach. *AV Communication Review*, 1973, **21** (*2*), 165-175.

STUFFLEBEAM, D.L. (Eds.) *Educational evaluation and decision making.* Illinois: F.E. Peacock, 1971.

WRIGHT, S. Path coefficients and path regressions: alternative or complementary concepts. *Biometrics*, 1960, **XVI**, 189-202.

YEE, A.H. The limits of scientific-economic-technological approaches and the search for perspective in education. *The Journal of Educational Research*, 1972, **66** (*1*), 19-29.

Research in Science Education, 1977, 7, 83-95.

FORMAL SCHOOLING AND THE ATTITUDE OF PAPUA NEW GUINEANS TOWARDS NATURAL PHENOMENA – A REPLICATION AND EXTENSION STUDY

M. N. Maddock

Synopsis

During 1974, field work for a replication and extension study was carried out, following on a pilot study conducted in 1972 to develop, refine and, in a comparative situation, field test an instrument to measure the attitude held by Papua New Guineans towards investigation, control and manipulation of natural phenomena. The extension study investigated a wider range of educational levels, ages of subjects and regional locations.

The findings of the original pilot study were supported in the replication phase, except that on this occasion, rural subjects scored higher than urban subjects.

In the extension phase, educational level was again found to be the major, significant variable in determining scores on the scale, while Regional differences were also found; Highland subjects having the highest mean. The level of cultural alienation between students and minimally educated villages, although somewhat smaller than in the original study, was still of significant magnitude.

No differences were found between results obtained by expatriates (white) and Papua New Guinean nationals doing interviewing in face to face interviews, but differences were obtained in group interviews.

Open-ended follow-up questioning revealed basically the same pattern of patronising depreciation of village culture by the educated subject as was found in 1972.

The pilot study

During 1972, fieldwork was carried out in a pilot study to develop, refine and field test an instrument to measure the attitude held by Papua New Guineans towards investigation, control and manipulation of natural phenomena. The full study is reported in Maddock (1973), while major aspects of the findings are reported in Volume 4 of this journal (Maddock, 1974). The study arose out of an interest by the researcher in evaluating attitudinal aims for the Papua New Guinea science curriculum.

In summary, a 26 item scale of high reliability called the Environmental Phenomena Attitude Scale was developed in four languages (English, Pidgin, Hiri Motu and Enga), administered under face to face interviewing conditions to a sample of 110 Papua New Guinean subjects consisting of grades 9 and 10 high school students matched by village subjects of about the same age and vernacular language, but with minimal or no formal schooling, and to a sample of over 300 high school students in group interview situations.

The scale was used in a comparative study in which it was found that students scored significantly higher than the villagers, but no regional, urban-rural or male-female differences were detected. The students predicted answers much lower on the scale for villager responses than their own, although the mean of the predicted scores was not as low as the mean actually obtained by the villagers.

In follow-up open ended interviews, the students were found to be patronising in their views of villager attitudes, attributing their own superiority to the fact that they were educated, but the villagers were ignorant. School science was often stated as a contributing factor to this.

The follow-up study

During 1974, on a grant provided by the Education Department of Papua New Guinea, fieldwork was conducted for a follow-up study, for which data analysis has been continuing since. A full report of this study can be found in Maddock 1977. The instrument was prepared in three more languages (Kua Nua, Toaripi and Motu) and administered to over 400 subjects in face to face situations, and over 1000 student subjects in group interviews.

The follow-up study incorporated a replication phase, which closely followed the original design, and an extension phase which included a wider range of levels of education, from zero formal schooling through to University level, a wider range of ages from about 11 years to middle aged adults, and a regional extension to the New Guinea Islands. In the original study, the interviewing was organised by the researcher, assisted by a number of expatriate (white) interviewers and used interpreters in administering The Papua New Guinean language versions. In the follow-up study a much bigger team of interviewers was used, including several Papua New Guinea nationals, the interviewers working independently and without interpreters. Prediction data was also obtained from the group interviews on this occasion.

A summary of the similarities and differences between the original and follow-up studies is set out in Table 1.

TABLE 1

Comparison of structure of pilot and follow-up studies

	Pilot Study	Follow-up Study	
		Replication	Extension
Languages	4: English, Pidgin Hiri Motu, Enga	7: additional: Kua Nua, Toaripi, Motu	7: additional: Kua Nua, Toaripi, Motu
Regions	2: Coastal, Highland	2: Coastal, Highland	3: Coastal, Highland N.G. Islands
Sample	*Face/Face* 52 (Grades 9-10) 48 (< Grade 4) Aged 11-24 *Group* 309 (Grades 8-10)	*Face/Face* 66 (Grades 9-10) 115 (< Grade 4)	*Face/Face* 42 (Grades 11+) 100 (Grades 7-8) 90 (Grades 5-6) Aged 11-40+ yrs. *Group* 1072 (Grades 7-12)
Interviewing Team	5 Male Expatriates with interpreters when required	4 Male Expatriates 2 Female Expatriates 8 Male PNG Nationals (no interpreters)	4 Male Expatriates 2 Female Expatriates 8 Male PNG Nationals
Predictions	*Face/Face* Students only, Grades 9-10	*Face/Face* Subjects Grades 9-10 education level *Group* Students Grades 9-10	*Face/Face* Subjects Grades 7-13+ education level *Group* Students Grades 7-12

Data analysis was conducted on the University of Newcastle's ICL 1904A computer using programmes devised by the researcher and various statistical analysis packages.

Instrument reliability

Satisfactory alpha reliability coefficients were obtained on the Environmental Phenomena Attitude Scale, which were very similar to those obtained in the original study. No alpha coefficients were determined specifically for the replication sample, but the whole sample alpha on the individual interviews was slightly lower than obtained on the 1972 results and the prediction alpha results obtained from the educated subjects were higher. The group interview alphas, obtained on student subjects only, were lower than those from the face to face interviews, as would be expected from a restricted range sample, but were higher than those from the 1972 study, with the Enga version still being the lowest. The new language versions produced very similar results to the English, Pidgin and Hiri Motu versions. The reliability data is summarized in Table 2 for the face to face interviews and in Table 3 for the group interviews.

TABLE 2

Reliability data: individual face/face interviews

Administration	Own Answer			Predictions		
	Educated	Min. Educated	Whole Sample	All Interviewers	PNG Interviewers	Expatriate Interviewers
1972 Pre-test	0.50	0.56	0.84	0.78	-	-
1972 Post-test	0.78	0.76	0.89	-	-	-
1974	0.70	0.63	0.71	0.84	0.80	0.88

TABLE 3

Reliability data: group interviews

Own Answers: Students only										
Administration	English-Pidgin			English-Hiri Motu				English-Motu		
1972 (Grades 8-10) Pre-test	0.45			0.53				0.40		
Post-test	0.48			0.59				0.54		
	English	Pidgin	H.Motu	Motu	Enga	Kua Nua	Toaripi	PNG Int	Exp.Int.	All
1974 (Grades 7-12)	0.64	0.62	0.62	0.66	0.42	0.61	0.58	0.63	0.66	0.65
Predictions										
1974 (Grades 7-12)	0.78	0.76	0.64	0.69	0.71	0.51	0.47	0.74	0.74	0.76

The replication

The replication sample consisted of 66 students in Grades 9 or 10 and a group of 115 minimally educated subjects of approximately equivalent ages and speaking the same vernacular as the student sample. The results obtained from the subjects' own responses in this sample were subjected to the same designs of 2x2x2 analysis of variance as in the original study (Maddock, 1974) to test coastal-highland, education level, urban-rural, and male-female effects.

The difference between the mean of the students' responses and the mean of their predictions was analysed for significance using a t-test.

Subjects' own responses

As in the pilot study, a highly significant education effect was obtained in which the student mean was considerably higher on the scale than the mean for the minimally educated subjects, and no coastal-highland or male-female effects were detected; although the mean for males was higher than the mean for females, almost reaching the 0.05 level of significance.

However, this study differed from the pilot study in that the urban-rural effect was statistically significant, with the mean for rural subjects being higher than the mean for urban subjects, particularly so in the Highland region. This dimension had been originally included in the analysis because it was felt that considerable urban experience, particularly regular contact with modern technology, might contribute to developing a more positive attitude as measured by the scale, especially in the area of application of investigation to technology.

A comparison of the results of analyses from the two studies is summarized in Table 4 and by comparing Figure 1 in this report with Figure 2 in the report on the pilot study (Maddock, 1974 p.126), and Figure 2 with Figure 3 in the same report (p.127), a graphical comparison can be obtained.

Prediction scores

Although there were statistically significant changes in means between the 1972 study and the replication phase (see Table 5), a very similar pattern of results was obtained in the replication phase. There was still a major, significant gap between student scores and the scores obtained from their predictions, but again the villager mean was not as low as the predicted mean.

The mean of scores for grades 9 and 10 students, obtained by face-to-face interviews, dropped from 69.2 to 66.6, the difference of 2.6 being significant at the 0.001 level (using a t-test for independent measures), representing 0.7 of a standard deviation, with a point biserial correlation of 0.32 (Cohen 1965, p.102).

The gap between the student and villager score had closed a little, with the 1974 mean for minimally-educated subjects being 0.7 scale points ($0.08\ \delta$) higher than in 1972, but this difference was not statistically significant.

The mean of the students' predictions for villager responses was 3.7 scale points higher ($0.4\ \delta$, $p < 0.05$), with a point biserial correlation of 0.2.

In 1972, the difference between the mean of the students' scores and the mean of their predictions was 21.0 (t 18.78, $p < 0.001$), with the two sets of scores correlating -0.26 ($p < 0.05$). In the replication phase of this study, the difference of 10.3 was substantially less.

TABLE 4

Comparison of analysis of variance results Pilot study and replication

Design and Means				
Total N 110 (170)	Coastal		Highland	
Students 69.2*(67.1)* Min.Ed. 55.6*(56.3)* Coastal 63.9*(61.1)* Highld. 62.9*(62.3)* Urban 62.6*(60.5)* Rural 64.4*(62.9)* Grand 63.3*(61.7)*	Urban	Rural	Urban	Rural
Students Cell Sample Size Mean Standard Dev.	17 *(11)* 68.6*(65.3)* 4.9*(3.95)*	13 *(29)* 69.2*(65.2)* 2.4*(4.2)*	13 *(9)* 70.2*(68.1)* 2.7*(4.6)*	19 *(13)* 69.2*(69.8)* 4.6*(2.5)*
Min. Educated Cell Sample Size Mean Standard Dev.	12 *(6)* 54.3*(56.0)* 4.7*(7.5)*	4 *(57)* 56.3*(57.9)* 3.9*(4.6)*	15 *(11)* 54.5*(52.5)* 7.9*(2.9)*	17 *(34)* 57.5*(58.6)* 6.0*(5.7)*

F Ratios, Epsilon and Square of Epsilon					
Effect	F	df F	P	E*	E^{2*}
Regional Education Urban-Rural	0.27*(1.59)* 157.71*(149.29)* 1.08*(7.00)*	1/102*(1/162)* 1/102*(1/162)* 1/102*(1/162)*	- 0.001*(0.001)* - *(0.01)*	- *(-)* 0.78*(0.69)* - *(.19)*	- *(-)* 0.61*(0.48)* - *(0.04)*
Sex	2.9	1/102	-	-	-

Note: 1974 Figures in Parenthesis
* See Cohen 1965, p.106.

TABLE 5

Differences between original study and replication phase

	1972	1974	d	d/δ	t	P<	df	r_{pb}
Students	69.2	66.6	2.6	0.70	3.6	0.001	116	0.32
Min. Ed.	55.6	56.3	−0.7	0.08	−1.39	ns	160	0.1
Predictions	47.6	51.3	−3.7	0.40	−2.3	0.05	112	0.2

It produced a t-test of 8.6, significant at the 0.001 level. In 1972 the difference between the prediction mean and the villager mean was 8.0 (t 5.88, $p < 0.001$), compared with the 1974 difference of 5.0 (t 4.0, $p < 0.001$). These results are summarized in Table 9. In cases where variances were shown to be significantly different, the t-test recommended by Yamane (1964, p.105) was used.

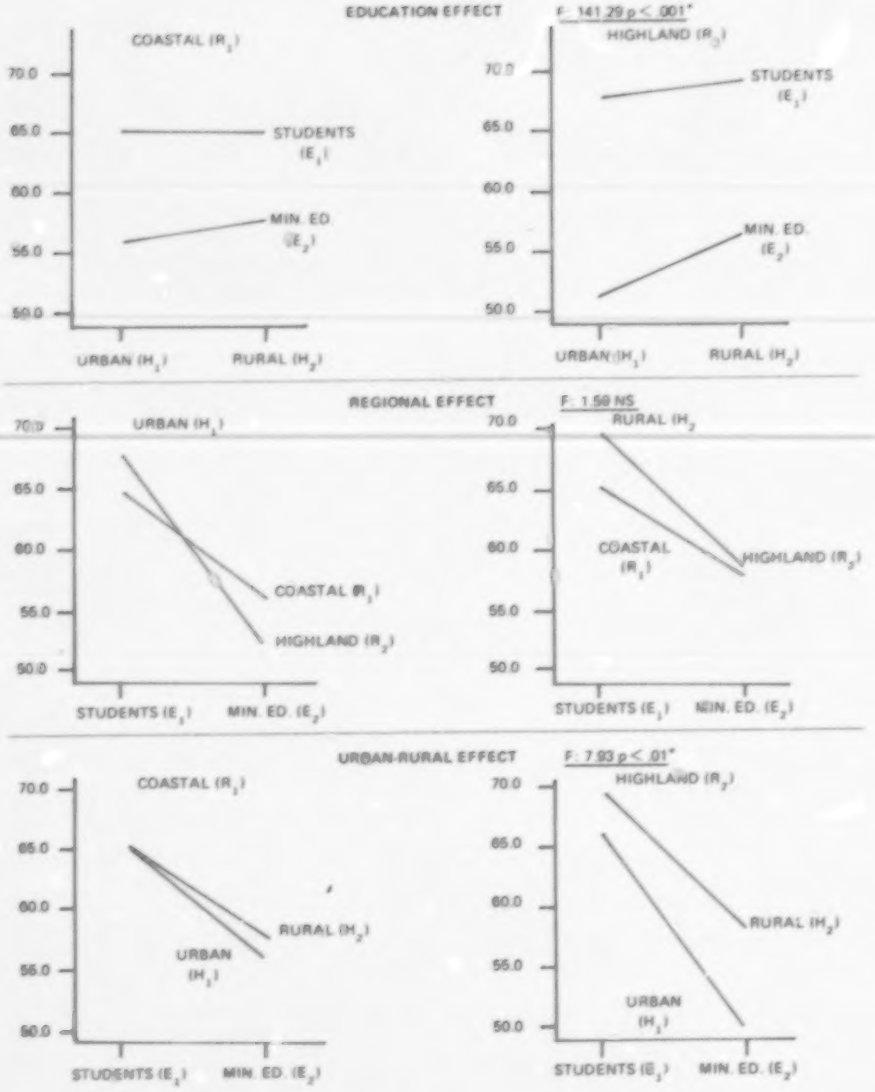


FIGURE 1

Plot of means, replication study (region, sex, education).

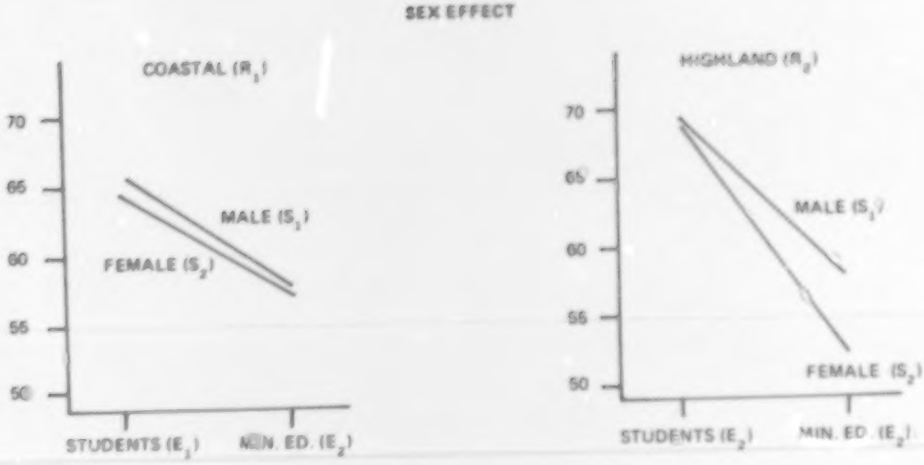


FIGURE 2

Plot of means, replication study (region, education, sex)

The extension

In the extended phase of the follow-up study, the regional dimension was expanded to include the New Guinea Islands (East New Britain and Duke of York) and the sample extended in age range and educational level. One hundred and forty two subjects at high school level or higher (42 with Grade 11+ and 100 with Grade 7 or 8) and 90 subjects who had left school, but had primary schooling at Grades 5 or 6 level, were added to the sample for face to face interviews, and a total of 1972 students ranging from Grades 7 to 12 were tested by group interviews in which the questions were read to a classroom group and respondents marked their answers on an answer sheet (see Table 1).

As in the original study, many difficulties, similar to those discussed in Maddock 1973, were encountered in obtaining suitable subjects in the various required categories.

Regional and educational level effects

In a 3 x 6 analysis of variance conducted on the total face to face interview sample, highly significant F values were found for both the Regional and Education effects.

The F value for the Regional effect was 8.01 ($p < 0.05$) regional differences accounting for about 3 percent of the variance (Cohen 1965, p.105).

When the Scheffé technique for post hoc multiple comparisons between means was applied (Kerlinger and Pedhazur, 1973, p.128-131), the mean for the Highland region (62.6) was found to be significantly higher than the Coastal mean (61.6) and the New Guinea Island mean (60.0), although the difference between Coastal and Island was not significant. In the replication section of the study, no regional difference between Coastal and Highland subjects was found. However, inspection of Figure 3 shows the Highland mean for level E_6 being much lower than the Coastal mean, thus accounting for the lack of significance in the reduced analysis.

The significant F for the Education effect provided an epsilon value of 0.49, representing about 24 percent of variance explicable in terms of differences in educational level.

There was a consistent superiority in the means for each education level over the previous one. Again applying the Scheffé technique (Kerlinger and Pedhazur, 1973, p.128-131), it was found that the differences were not significant between E_1 and E_2, E_2 and E_3, E_4 and E_5, E_5 and E_6, but the difference between E_3 and E_4 was significant. However the differences between E_1 and E_3 and E_3 and E_5 and E_4 and E_6 were significant. The difference between the mean of E_1, E_2 and E_3 combined (high school education or higher) and E_4, E_5 and E_6 (primary schooling or nil) was strongly significant. Thus the educational level of the subject was a major factor in determining the score obtained on the scale. These results are summarized in Table 6, and illustrated graphically in Figure 3.

TABLE 6

Summary of analysis of variance
F values, epsilons and mean differences.
Whole sample, face/face interviews

Effect	F	P	ε	ε^2	Variable	Mean	Mean Combined Scores
Regional	8.01 df (2/395)	<0.001	0.19	0.03	Highland Coastal Island	63.6 61.6 60.6	
Educational	26.45 df (5/395)	<0.001	0.49	0.24	E_1 (Grades 11+) E_2 (Grades 9,10) E_3 (Grades 7,8) E_4 (Grades 5,6) E_5 (Grades 3,4) E_6 (Grades 0,1,2)	67.9 65.6 62.9 59.5 58.0 56.3	64.8 58.0

*Significant Difference, Scheffe test, $p < 0.05$ N = 413.

Prediction data

All educated subjects, both in individual and group interviews, were asked to predict villager answers, and the pattern of responses closely followed that of the "own answer"-prediction pattern of the original study, obtained only on face to face interviews, and of the replication phase; although the differences were not quite as large (see Table 7).

In the 1972 study, a significant negative correlation of 0.26 was obtained when the student scores were correlated with their predictions, indicating a distinct tendency for students to predict lower responses for their village counterparts, the higher they scored on the scale.

In the 1974 study, for face to face interviews, between the scores of educated subjects (Grade 7 to 11+) and their predictions the correlation was -0.21, and for group interviews the correlation was -0.36, both significant at the 0.001 level. The differences of 13.4 for the face to face data and 15.8 for the group data were both highly significant. Thus two years later, over a much wider sample, the tendency for students to more strongly depreciate their village

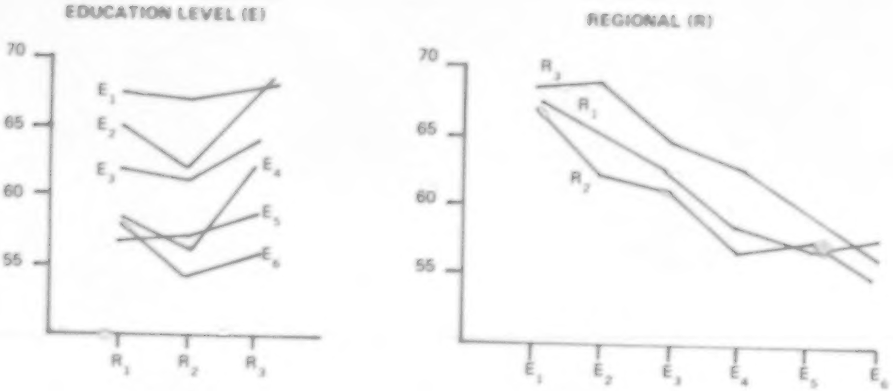


FIGURE 3

Plot of means, whole sample, face to face interviews, subjects own answers, Region by Education level

counterparts the higher they scored on the EPAS scale was still present, although the gap was not as intense as in the 1972 study.

As in the original study, the mean of the scores actually obtained by the minimally educated subjects was not as low as the mean of the predicted scores, although the gap again was somewhat smaller than in 1972.

Figure 4 gives a graphical summary of the score differences for the original study and the replication and extension phases.

TABLE 7

Student responses, predictions and villager scores
1972-1974

	1972		1974					
	Face/Face		Replication Face/Face		Face/Face		Group	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Students (S)	69.2	3.9	66.6	3.9	64.3	6.1	64.7	5.8
Min. Ed. (M)	55.6	0.0	56.3	5.02	58.5	6.3	-	-
Predictions (P)	47.6	8.0	51.3	8.8	50.9	8.5	48.9	7.6
S-P	21.6		10.3		13.4		15.8	
$t_{(S-P)}$	18.78		8.6		18.38		53.7	
$p <$	0.001		0.001		0.001		0.001	
df	85		93		349		1923	
M-P	8.0		5.0		7.6		-	
$t_{(M-P)}$	5.88		4.0		10.31		-	
$p <$	0.001		0.001		0.001		-	
df	129		108		360			

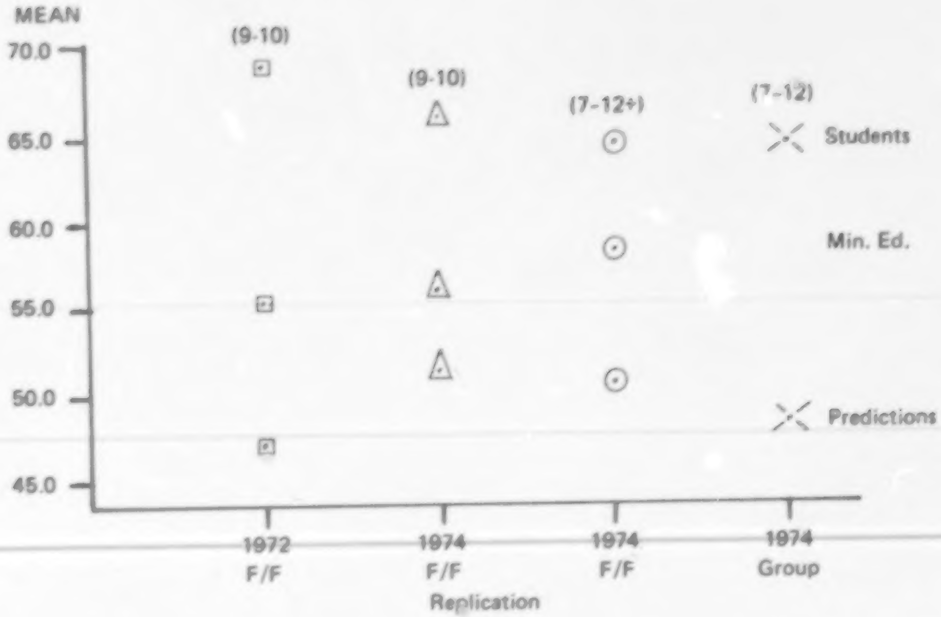


FIGURE 4

Student, minimally educated and student prediction means compared 1972-74 face to face – 1974 group.

Interviewer effect

In the original pilot study, a check on interview reliability was carried out which did not reveal any significant variation between interviewers. However, Papua New Guinean nationals were involved only as interpreters and the expatriate influence was always evident. An important phase of the follow-up study was to check whether Papua New Guineans obtained different results from Expatriate interviewers, as well as providing training to interviewing and research techniques for a number of Papua New Guinean nationals, who were involved in testing with both the English and the Papua New Guinean language versions, and made independent tours of regional areas to carry out the field work.

There were no differences between results obtained by expatriate interviewers and those obtained by Papua New Guinean interviewers in face to face interviews, either for the whole sample including both students and villager subjects, or for the prediction obtained from the students, although significant differences were found in the case of the group interviews (see Table 8).

In the case of the group interviews (with students only), the mean of scores obtained by expatriate interviewers was 2 scale points (about 0.35 δ) above that obtained by Papua New Guineans, significant at the 0.001 level, and accounting for about 3% of the variance (Cohen 1965, p.102). However, in the case of predictions obtained by group interviews, the

TABLE 8

Interviewer effect

	Own Answer		Predictions	
	Face/Face	Group	Face/Face	Group
Expatriate Mean (E)	62.6	65.8	50.7	46.8
PNG Mean (P)	62.4	63.8	51.0	50.7
N (EXP., PNG)	127,236	481,591	63,102	468,524
$d_{(E\text{-}P)}$	0.2	2.0	0.3	3.9
δ (EXP., PNG)	7.1,6.4	5.6,5.7	10.5,8.6	7.1,7.4
d/δ (Approx.)	0.03	0.35	0.03	0.54
t	0.84	5.8	0.20	8.4
$p <$	n.s.	0.001	n.s.	0.001
df	361	1070	163	990
r_{pb}	–	0.17	–	0.26
r_{pb}^2	–	0.30	–	0.70

students predicted significantly higher responses to Papua New Guinean interviewers than to Expatriate interviewers, with a difference of 3.9 scale points (about $0.54\ \delta$). About 7% of the variance is accounted for by the interviewer effect, compared with about 3% in the case of the students' own answers.

Open-ended questioning

As in the 1972 study, open-ended follow-up questioning was carried out with educated subjects to determine reasons for choosing specific alternatives and the predictions they made for villagers' answers; and to explore their knowledge of traditional village culture as it related to the questions used in the instrument. The interviewing was done in all 7 languages, all interviewers were involved and the interviews recorded on tape, then translated and transcribed. Analysis of this information is still proceeding.

To date, a very similar pattern of responses to those obtained in 1972 is emerging. When asked why they had predicted a different answer for the villager response to their own choice, the common trend of the replies was to depreciate villagers as ignorant, without schooling, and intent on adhering to traditional ways, whereas the student had been educated and knew better. Science lessons at school were again often mentioned as a contributing factor to the students' views.

However there seems to be a significant improvement in the knowledge of traditional village legends and practices since the 1972 study, where very few of the respondents were able to provide information in this area.

Conclusions

The Environmental Phenomena Attitude Scale produced satisfactory reliability coefficients of the same order as those obtained in the original 1972 study, and the new language versions (Motu, Kua Nua and Toaripi) produced results highly comparable with those from the original versions.

In the replication phase, using a similar sample of Grades 9 and 10 students and a matching villager group with minimal education, a very similar pattern of results was obtained, with Educational Level a significant factor, and Coastal-Highland and Male-Female effects non significant. The replication study differed in that a significant Urban-Rural effect was obtained, with rural subjects scoring higher than urban subjects.

The mean of the students' score was significantly lower and the mean of their predictions for villager answers significantly higher than in 1972, but the mean for the minimally educated subjects was about the same. However, the pattern of differences between these three was basically the same, with the students predicting villager responses considerably lower on the scale than their own, but not as low as those actually obtained from villager respondents.

In the extended study, Educational Level was the major factor in determining attitude as measured by the scale, with the mean score rising for each educational level, with significant differences between high school educated and primary school or less educated subjects, between Grades 5 and 6 and Grades 7 and 8 educated subjects, and between a level of education and one two levels above it.

When the additional subjects were added to the replication sample, a Regional Effect emerged in which Highland subjects scored significantly higher than Coastal or New Guinea Island subjects.

Similar negative correlations were obtained between the scores of educated subjects and the scores obtained from the predictions of villager responses, as in the original study, with subjects scoring highest on the EPAS scale tending to predict responses lowest on the scale. The pattern of differences between the scores of students, their predictions and scores actually obtained from minimally educated subjects, was of the same order, but somewhat reduced. As in 1972, Villagers did not produce scores as low on the scale as those obtained from the predictions by students.

There were no differences between results obtained by expatriate and Papua New Guinean national interviewers, working independently, except in the group interview situation in which the mean of EPAS scores obtained by expatriates was slightly, but significantly lower than that obtained by Papua New Guineans and subjects made lower predictions for village scores to expatriate interviewers than to Papua New Guinean interviewers.

Open-ended follow-up questioning revealed a very similar pattern of replies to that received in 1972 with students depreciating village culture as one of ignorance, although they seemed much more knowledgeable about traditional beliefs and practices than was found in the original study.

Implications

This study provides evidence that, in respect of attitude to the environment as measured by the Environmental Phenomena Attitude Scale, formal schooling is having an alienating effect which has remained strong over the period 1972-1974. Such alienating effects can provide potential for conflict between highly educated and minimally educated subjects. Papua New Guinean educators need to examine this aspect of formal schooling very closely, to decide whether in fact that it is a price they are prepared to pay for formal schooling, or whether steps need to be taken to alleviate such alienation.

Acknowledgments

The study was undertaken on a grant from the Department of Education in Papua New Guinea. A study of the magnitude and complexity of this one in the Papua New Guinea situation could not have been carried out without the help of a large number of people, too numerous to name here, but thanks are extended to all who helped, particularly the interviewing team.

References

COHEN, J. Some statistical issues in psychological research. In Wolman, B.B. (Ed.), *Handbook of Clinical Psychology*. N.Y.: McGraw-Hill, 1965.

KERLINGER, F.N. & PEDHAZUR, E.J. *Multiple regression in behavioural research.* N.Y.: Holt, Rinehart & Winston, 1973.

MADDOCK, M.N. A pilot study to develop, refine and field test an instrument to measure the attitude of Papua New Guineans towards the investigation, control and manipulation of natural phenomena. Unpublished Ph.D. dissertation, Florida State University, U.S.A., 1973.

MADDOCK, M.N. Measuring attitudes held by high school students and minimally educated villagers in Papua New Guinea – An instrument development study. *Research in Science Education*, 1974, **4**, 113-129.

MADDOCK, M.N. The attitude of Papua New Guineans towards natural phenomena, a replication and extension study. Report to the Education Department of Papua New Guinea, The University of Newcastle, 1977.

YAMANE, T. *Statistics: an introductory analysis.* N.Y.: Harper & Row, 1964.

Research in Science Education, 1977, 7, 97-103.

SUGGESTIONS FROM CHINA ON THE "RELEVANCE" OF SCIENCE INSTRUCTION

Frederick J. Thomas

All in all, science education seems to work fairly well for the middle class of developed nations. There are problems, of course, but at least in terms of careers and manpower needs, science education is relatively successful in helping prepare the doctors, engineers, laboratory technicians, television repairmen, air traffic controllers and other people whose use of science is essential to the functioning of our societies.

Science education seems to be less successful with certain cultural minorities within the developed nations. In the United States, for example, few Black children or American Indian children gain enough from science instruction to allow their entry into the more prestigious science-related jobs. There have been several programs intended to correct this imbalance, but the results have generally been disappointing.

The most distressing failures of science education, however, have probably occurred in the Third World. Many of the poorer nations which became independent after World War II sought to use science instruction as a major driving force behind "modernization". Often relying extensively upon imported textbooks and foreign teachers, these nations invested heavily in science teaching, assuming that an increased knowledge of science would help bring about increased agricultural and industrial production, decreased birth rates, and other characteristics of "modernization". Results have fallen far short of expectations, and a part of the blame has been placed on the "irrelevance" of foreign models of science education. Perhaps the books, teaching procedures, institutional arrangements and other aspects of science education which work for the middle class of America, Britain and Australia should not have been expected to work in other national and cultural environments. At present, many nations are devoting a great deal of effort to developing original, "more relevant" forms of science education (Maybury, 1975; Poats, 1972). There have also been important attempts to develop "more relevant" science instruction for cultural minorities within developed nations, as described, for example, by Kalra (1975) and Snow (1972).

Unfortunately, it is difficult to see just how science instruction can be made more "relevant" to particular nations or cultures. Many of the local development projects seem to produce little more than second-rate imitations of what could have been purchased less expensively abroad. When the local materials *are* of high quality, they often seem exportable. Perhaps what is "relevant" for students in one nation or culture is almost inevitably "relevant" to students in other nations and cultures.

Aside from the more absurd cases of irrelevance (as when African students worked with plants imported from Europe for their biology exams), there may be little need for concern. Science education has failed to overcome national and cultural inequalities, but there are certainly many complex causes for the inequalities. Perhaps it is unreasonable to expect minority and Third World students to gain much more than at present from science instruction.

In support of the argument against an excessive concern for cultural or national "relevance", it should be noted that science is one of the more universal school subjects. The laws of physics and the principles of sexual reproduction are quite the same in Australia as in the United States, Britain, Ghana, Brazil, India, the Soviet Union or China. There is little doubt that scientists from all these countries manage to communicate and co-operate far more easily than do politicians, businessmen, philosophers or theologians.

More than any other nation, the People's Republic of China has been cited in the Western literature as an example of a nation which teaches and uses science in a unique way (Dean, 1972; Rifkin, 1975; Sidel, 1973; Suttmeier, 1974; Swetz, 1974). Although some authors believe the Chinese have damaged science in making it "relevant", many claim the Chinese have demonstrated the need for every nation (whether poor or developed) to devise its own, unique approach. The argument, however, has been characterized by a shortage of hard data. No extensive research conducted within China has been published, and only a few studies (notably Ridley, Godwin and Doolin, 1971) based on systematic examination of Chinese instructional materials are available.

In an effort to determine how – if at all – the Chinese had made their science education more "relevant", this author (Thomas, 1976) examined a range of printed materials which the Chinese claim to have used extensively in science education during or since the Cultural Revolution. The study centred around education in meteorology for persons other than professional meteorologists. The materials examined included a series of science readers (*One Hundred Thousand Whys*), newspaper articles on meteorology in the Peking daily *Kuang-ming Jih-pao*, and several essays by Mao Tse-tung. It was assumed that the newspaper articles were intended to educate non-scientists (particularly people in positions of political or economic authority) about how meteorology functioned – or should function – in China. The Mao essays have been studied intensely by virtually everyone in China, often in relation to science education. The results of that study are summarized here as the basis for some suggestions to other cultural groups or nations (including Australia) as to how their science instruction might be made more "relevant". It is not suggested here that any group follow China in the way many have followed Britain or the United States. The aim is instead to help identify the specific variables which the Chinese seem to have found important, and to consider how those variables might be important to others.

Four aspects of "relevance" were assumed to be important, and not investigated extensively. These were (1) language, (2) selection of examples, (3) mechanical structure, and (4) selection of topics or problems. Obviously, students need to be able to understand science instructional materials, and this requires either the use of "relevant" language or prior instruction in the language used. It should also be obvious that many examples should be drawn from the students' actual environment. Problems sometimes arise if materials written for one mechanical structure (say, the pattern of five fifty-minute periods) are used in other structures (such as three-hour sessions, once per week for adults). The selection of topics should also be "relevant", as when North American texts might emphasise the health problems of obesity while West African texts might place more emphasis on malaria.

Unfortunately, a great deal of science instruction is not "relevant" in the four ways suggested above, but it is still important to consider whether these four variables are sufficient. If these four variables really cover all that is required to make instruction "relevant", one conclusion might be that most local or national instructional development projects are a waste of time. Except for some topics of only local interest, "relevance" in these four ways could be achieved through the careful selection and adaptation of foreign texts and other materials.

The point of examining Chinese science instructional materials was to see if they had gone beyond these four aspects of "relevance" to produce an education which was somehow better for China than that which could have been obtained by selecting and adapting previously available foreign materials. The conclusion was that the Chinese probably had made considerable progress in several other aspects of "relevance".

To understand how meteorology and meteorological education function in China, it is necessary to consider very briefly how the structure of the national weather service relates to China's economic and social structure. Under the revolutionary rhetoric, China's meteorological service seems fairly conventional. It is centred in Peking, with branches in each province. Meteorologists use surface observations, high altitude balloons, weather radar, satellite cloud pictures and other contemporary tools to issue short-term, intermediate and long-term forecasts. During the last 10 or 15 years, the service seems to have been expanded considerably, to provide stations in the great majority of China's 2000 counties. These county stations make weather observations, and may issue "supplementary forecasts", based largely on the more general forecasts produced at higher levels.

Within a county, however, there might be 30 or more communes, each divided into several production brigades. It is apparently at the level of production brigades where meteorological information would most likely be used in making agricultural decisions – for example, when to plant or harvest a crop. It should be noted that many such decisions in China are made at a much higher level than that of individual farmers, as would be the case in most of the United States, Australia, or in most agricultural systems of the Third World.

In addition to extending the meteorological service to the level of counties, the Chinese have encouraged the establishment of "meteorological points", run by amateurs or part-time workers. There are reported to be 10,000 such posts ("Meteorological Posts", 1974), most organized by individual communes or production brigades. This would seem to indicate the establishment of posts on only about 15% of all communes. Most of the posts seem to be run by "educated youth" – young people with some secondary education who have been encouraged (some would say "forced") to settle in rural areas. Most would seem to have had little or no training in meteorology, except perhaps the type of training provided by books like *One Hundred Thousand Whys*.

It should be noted that almost all of these "educated youth" would have been born since the Chinese Revolution of 1949, and that many were raised in cities. Among the farmers on the commune, there are certain to be many who have spent a good deal of time watching the weather. There may also be older men and women on the commune with vivid memories from before 1949 of a hailstorm, flood or drought which destroyed a year's harvest and led to massive starvation and suffering.

Between 1966 and 1975, twelve million "school graduates" were reported to have settled in rural areas ("Twelve Million", 1976). This would seem to indicate roughly one school graduate (including graduates of junior secondary school) for every 50 rural residents. What sort of science education would be most "relevant" for these young people? What sort of education would be most "relevant" for the adults – ranging from commune administrators to illiterate peasants – with whom the young people must work?

In several ways, the Chinese seem to have concluded that what is "relevant" for them is very similar to previously published Western science materials. *One Hundred Thousand Whys* includes descriptions of the Chinese meteorological service, explanations of Chinese weather phenomena (cold waves, typhoons, etc.), instructions for using simple weather instruments (barometers, human-hair hygrometers, automatic rain guages, etc.) and discussions of important local weather variations (such as the formation of frost in low places). All of this seems quite useful and "relevant", particularly to an "educated youth" assigned to a commune weather post, but only the language and some of the examples are uniquely Chinese.

Some of the more unusual sections of *One Hundred Thousand Whys* deal with peasant sayings. One saying, for example, is that, "If there is a rainbow in the east it will be sunny; if there is a rainbow in the west it will be rainy". The truth of the saying is explained in terms of

China's prevailing winds, which blow a storm in the east away from the observer while a storm in the west is blown towards the observer. Other sections discuss the use of animals in short-term forecasting. For example, "When the toads come out from the holes it is going to rain" is explained in terms of toads' sensitivity to humidity. Few Western science texts give much emphasis or credulity to peasant sayings, but Humphreys (1923) and Smith (1958) are at least two reputable science writers who considered very similar sections to be "relevant" for American and British readers.

Chinese newspaper articles may show a more unique sort of "relevance". For example, "The Contributions of Shen Kuo, a Legalist, to Meteorology" by the Atmospheric Physics Research Centre of the Chinese Academy of Science (1975) is part of the broad "anti-Confucist movement" in China. This movement involves a deep reinterpretation of the history of science in China, not only to show that China has had important scientists for many centuries, but also to blame China's scientific and technological backwardness on the followers of Confucius. Other newspaper articles (Yi Wu County Weather Station, 1975; Shih Ping Brigade, 1975) go much further than *One Hundred Thousand Whys* in illustrating the processes by which peasant sayings can be "summed up" and tested, and combined with "foreign tools" to yield improved forecasting methods.

It is very difficult to make sense of the Cultural Revolution in China. In 1966, schools closed, science journals stopped publishing, and millions of Red Guards roamed over China, carrying the writings of Chairman Mao. Among the most widely studied works were the "Three Constantly Read Articles", which apparently reached many people previously untouched by anything resembling formal education. The three articles had been written by Mao many years before, during the fighting with Japan and the Kuomintang. The primary theme of all three is selfless service and the need for long, hard work.

Although it may seem absurd to accept Chinese claims that the Mao articles made a positive contribution to science education – perhaps a contribution which justified the closing of schools for several years – the claim does deserve serious consideration. The articles are not "science instructional materials" in any obvious way, but would more properly be called "philosophical" or "social" or "political" writings. Still, they include some important bits of science. In a simple, direct way, "Serve the People" says that all authorities – including meteorologists and other scientists and technicians – ought to serve the people. In a country where the few Westernized scientists had always been very much of an elite, this simple message, spread throughout the nation with the full authority of Chairman Mao, might have been extremely important.

Strangely, one of the three articles *"In Memory of Norman Bethune"*, is about a Canadian Communist and doctor who served and died in China during the Revolution. One important message of this article would seem to be that "foreign" science can be a very good way of carrying on the Revolution. It should be remembered that medicine in China has traditionally not been a high-status profession, in line with the general Confucian attitude that any hands-on science is beneath the dignity of a scholar.

The third article, *"The Foolish Old Man Who Removed the Mountains"*, cites an ancient Chinese fable about the Foolish Old Man of North Mountain, who set out with his sons to dig up two mountains blocking his way to the south. The Wise Old Man said he was silly, but the Foolish Old Man replied that the mountains could not grow, and if his sons and grandsons, and their sons and grandsons, kept digging, the mountains could be cleared away. In the original fable, God was moved by the Old Man and sent two angels to carry away the mountains. In the Maoist reinterpretation, the two mountains are imperialism and feudalism, and "Our God is none other than the masses of the Chinese people".

Another unique but important component of Chinese science instruction is provided by Mao's longer essays, "On Practice" and "On Contradiction". These essays have been cited as providing the correct methodology for thinking about scientific and technical problems. Put very briefly, the main points of the Maoist methodology are: (1) All real knowledge is based upon experience, whether personal experience or the experience of others, (2) knowledge must be constantly deepened, through a cycle of "practice, knowledge, again practice, and again knowledge", (3) both theory and practice are essential – theories "sum up" practical experience, and practice is the way of testing and improving theories, (4) to understand something, it is necessary to study both internal and external factors, (5) external factors are important conditions of change, but the primary cause of change is always internal, and (6) there is always one internal "contradiction" which predominates.

Particularly with the last two points, Mao's methodology is quite dogmatic and probably antiscientific. Scientists in China seem, fortunately, to have done a rather good job of resisting or side-stepping the imposition of the methodology. Restricting our attention, however, to less well educated workers (such as part-time workers at a brigade meteorological post), might the Maoist essays help promote a more scientific way of thinking? They would probably be at least as effective as the parts of "The Scientific Method" which Western students have sometimes recited in science class. Particularly since the Maoist essays are the corner-stone of a vast, integrated effort to change the way people think, their impact must surely have been far greater than any method advocated only within science class.

The Chinese seem to have done a rather good job of making their science instructional materials "relevant" in the four ways mentioned above – language, selection of examples, mechanical structure, and selection of topics or problems. They seem to have been particularly successful in adapting some fairly conventional topics (such as rain gauges or the social value of science) for nonschool instruction of both young people and adults.

The Chinese have felt it necessary, however, to go further in search of "relevance", and it is these additional steps which may provide some important suggestions for other nations and cultures.

The Chinese are teaching a new interpretation of their indigenous science. American science teachers have long brought out George Washington Carver and Marie Curie to prove to Blacks and women that they can be a part of science, but the Chinese have gone much further towards explaining how science functioned in ancient China – and why it did not progress further. The Chinese intepretation may be too dogmatic, but at least they are dealing with the issue. Perhaps science education could be made more "relevant" to Greeks, Turks, Aboriginals and other Australian minorities if it explored both the successes and failures of science in these cultures.

In addition to discussing the history of science in China, Chinese instructional materials place a very strong emphasis on developing respect for the science-related knowledge of nonscientists. This emphasis might be too great if students are led to accept blindly the truth of all peasant sayings, but the reverse situation is equally wrong. Perhaps American and British and Australian science texts go too far towards teaching that knowledge about the world comes only from "experts". Perhaps the daughter of a Chinese fisherman and the son of an American plumber would both be better off if they were encouraged to combine their parents' knowledge with the knowledge of their science teachers.

The Chinese are working quite hard to reduce the elitism of science, both by teaching peasants that science should "serve the people" and by "re-educating" scientists. Their science teaching incorporates their social values about the social status of scientists and about the sociology of science. They readily admit that they make science political, when they demand that science serve the workers. The Chinese go on to claim, however, that *all* science education is deeply political, and that in America and Australia it serves capitalism, by teaching such values as individual competition.

As has often been stated, science is not just a body of facts but also a way of thinking. In Chinese education, the "scientific way of thinking" is presented as being identical with "Mao Tse-tung Thought". This one way of thinking is said to apply equally to science, to philosophy, to politics, and to all other areas. In the West, the role of the "scientific method" taught in science classes is more vague. Some would say that the method should be applied to Parlimentary votes, to legal decisions, to religious questions, and to many other areas. Others would say that the "scientific method" is valid only for certain types of questions, and must not be injected into the domain of philosophers, theologians and others. In any case, the issue might need to be explored further if the scientific way of thinking is to be made more "relevant" to students.

More than any other nation, China has broken with Western patterns in search of a more "relevant" science education. Few nations will want to adopt Chinese textbooks, even in translation, but the questions they raise may be of importance to everyone.

References

ATMOSPHERIC PHYSICS RESEARCH CENTRE OF THE CHINESE ACADEMY OF SCIENCE. The contributions of Shen Kuo, a legalist, to meteorology. In F.J. Thomas, An exploratory study on the cultural relevance of science instructional materials. Unpublished doctoral dissertation, Indiana University, 1976. (Reprinted from *Kuang-ming Jih-pao*, August 26, 1975, p.3.)

DEAN, G. China's technological development. *New Scientist*, 1972, 54, 371-373.

HUMPHREYS, W.J. *Weather proverbs and paradoxes*. Baltimore: Williams & Wilkins, 1923.

KALRA, R.M. Science taught with a focus on values. *Journal of American Indian Education*, 1975, 14 (*2*), 21-25.

MAO TSE-TUNG. On contradiction. In *Selected works of Mao Tse-tung* (Vol. 1). Peking: Foreign Languages Press, 1967. (Originally given as a lecture at the Anti-Japanese Military and Political College in Yenan, August 1937.)

MAO TSE-TUNG. On practice. In *Selected works of Mao Tse-tung* (Vol. 1). Peking: Foreign Languages Press, 1967. (Originally given as a lecture at the Anti-Japanese Military and Political College in Yenan, July 1937.)

MAO TSE-TUNG. In memory of Norman Bethune. In *Selected works of Mao Tse-tung* (Vol. 2). Peking: Foreign Languages Press, 1967. (Originally written December 21, 1939.)

MAO TSE-TUNG. Serve the people. In *Selected works of Mao Tse-tung* (Vol. 3). Peking: Foreign Languages Press, 1967. (Originally a speech delivered at a memorial meeting, September 8, 1944.)

MAO TSE-TUNG. The foolish old man who removed the mountains. In *Selected works of Mao Tse-tung* (Vol. 3). Peking: Foreign Languages Press, 1967. (Originally a speech delivered at the Seventh National Congress of the Communist Party of China, June 11, 1945).

MAYBURY, R.H. *Technical assistance and innovation in science education*. New York: John Wiley & Sons, 1975.

Meteorological posts in the countryside. *Peking Review*, December 6, 1974, p. 23.

One hundred thousand whys. (12 vols.). Shanghai: People's Publisher, 1970-1973.

POATS, R.M. *Technology for developing nations.* Washington: Brookings Institute, 1972.

RIDLEY, C.P., GODWIN, P.H.B., & DOOLIN, D.J. *The making of a model citizen in communist China.* Stanford, Calif.: Hoover Institute, 1971.

RIFKIN, S.B. The Chinese model for science and technology: its relevance for other developing countries. *Development and Change*, 1975, 6, 23-40.

SHIH PING BRIGADE. Determined to control the weather properly. In F.J. Thomas, An exploratory study on the cultural relevance of science instructional materials. Unpublished doctoral dissertation, Indiana University, 1976. (Reprinted from *Kuangming Jih-pao*, July 15, 1975, p.3).

SIDEL, V.W. Medicine and public health. *Proceedings of the Academy of Political Science*, 1973, 31, 110-120.

SMITH, L.P. *Farming weather.* London: Thomas Nelson & Sons, 1958.

SNOW, A.J. Ethno-science in American Indian education. *Science Teachers*, 1972, 39 (7), 30-32.

SUTTMEIER, R.P. *Research and revolution: science policy and societal change in China.* Lexington, Mass.: Heath, 1974.

SWETZ, F.J. *Mathematics education in China: its growth and development.* Cambridge, Mass.: MIT Press, 1974.

THOMAS, F.J. An exploratory study on the cultural relevance of science instructional materials: the People's Republic of China. Unpublished doctoral dissertation, Indiana University, 1976.

Twelve million school graduates settle in the countryside. *Peking Review*, January 9, 1976, pp.11-13.

YI WU COUNTY WEATHER STATION. Depend on the mass of people to organize good weather stations. In F.J. Thomas, An exploratory study on the cultural relevance of science instructional materials. Unpublished doctoral dissertation, Indiana University, 1976. (Reprinted from

Research in Science Education, 1977, 7, 105-111.

A STUDENT'S-EYE VIEW OF THE BIOLOGY CLASSROOM: TEACHER, ACTIVITIES AND OUTCOMES

Effie D. Best

Introduction

During 1973 and 1974 I collected a wealth of data from South Australian Matriculation Biology classes. One of the purposes of collecting these data was to give feedback to teachers about how their students saw classroom activities and the outcomes of learning biology. Another was to see whether the data revealed any patterns of relationships between activities and outcomes as seen from the student's viewpoint. This paper reports some relationships found in the 1974 data which give some support to the value of student-initiated inquiry as a component of biology teaching. The data also indicate that the students' perceptions of teacher-pupil relationships are related to the way they perceive other aspects of the classroom.

Teachers and students involved

All matriculation biology teachers were asked whether they wished to be involved in the project in both 1973 and 1974. The 1974 teachers included many (but not all) of those who participated in 1973, and a few who had not participated before. I asked them to administer two instruments to their classes at the end of the year (except in very small classes, where only one of the two was administered). I sent teachers a package with the *Biology Activities Checklist: Student Perceptions*, (BAC:SP) interleaved with the *Student Outcome Checklist: Student Perceptions*, (SOC:SP) and asked them to hand these out as they came out of the package. Packages started randomly with either the BAC:SP or the SOC:SP. So half the class answered each instrument. All students in some classes also answered the *Teacher-Pupil Relationship Scale* (if the teacher chose to include this).

Instruments used

Biology Activities Checklist: Student Perceptions (BAC:SP)

I have already described stages in the development of this 60 item instrument and of attempts to subdivide it into scales (Best, 1975). The data collected at the end of 1974 (from 719 students in 60 classrooms) showed factor patterns very similar to those shown by the data collected at the end of 1973. I made minor changes to some subscales (mainly by deleting items) and combined two of the 1973 subscales (then called "Demand for Interpretation" and "Questioning (Laboratory)") into a single scale called "Investigation". Table 1 shows representative items from each subscale, together with the value of Cronbach's alpha calculated from the 1974 data giving an indication of the internal consistency of the subscale and the F value from a one-way analysis of variance (giving some idea of how the subscale discriminates among classes).

Twelve items did not fit clearly into any subscale. I will supply a list of items and subscales to anyone who wants it.

TABLE 1

Subscales of the Biology Activities Report: Student Perceptions

Scale	Representative items (Students responded "yes" or "no")	No. of items	Cronbach's alpha	F value (59, 659)
Investigation	12. We often think about the evidence that is behind statements made in the textbook. 42. We have a chance to analyse the conclusions that we have drawn from our practical exercises.	17	.76	4.3
Student initiative	32. We sometimes design our own experience to seek answers to questions that puzzle us. 41. We sometimes do some additional experimenting of our own when we have finished the set practical exercise.	9	.65	6.0
Demand for memorization	1. We are expected to write down and memorize what the teacher tells us. 20. Our tests often ask us to write out definitions of terms.	5	.38	2.8
"Don't ask what it means, just get on with it"	26. We usually understand the purpose of a practical exercise before we start it. (scored in reverse) 58. We often spend so long on a topic that we are bored.	11	.63	5.0
"Teacher knows best"	8. If there is a disagreement among students during a discussion, the teacher usually tells us the answer. 40. Our teacher usually tells us what the results of our experiments should be.	4	.44	2.8
Free discussion	4. If I don't agree with what the teacher says, I feel free to say so. 55. In discussions we feel free to express our opinions and ideas even if we are not sure if they are right.	2	.56	1.5

Student Outcomes Checklist: Student Perceptions (SOC:SP)

In 1974, 716 students from 62 classes responded to the 60-item checklist. The factors that emerged differed slightly from the 1973 factors, mainly because a number of items concerned with man and environment grouped together in 1974. I scored these as a separate scale (Environmental awareness). Most of the other scales were similar to the 1973 scales with the rearrangements one would expect to result from loss of items from other scales to the Environmental awareness scale. Table 2 shows representative items from each subscale, with Cronbach's alpha calculated from the 1974 data and the F value from a one-way analysis of variance with class as the independent variable.

Although the F values show that there are significant ($p < .05$) differences among classes, they are quite low, particularly compared with the F values for the BAC:SP subscales. It is not surprising that individual differences among students contribute more to their perceptions of the outcomes of learning biology than to their perceptions of what happens in the classroom. This means, however, that the analyses of class means reported later in this paper give only part of the classroom picture as viewed by individual students.

I will supply a list of items and subscales (including 13 items that do not fit into subscales) to anyone interested. It seems to make sense to use a total score for the outcome checklist, including unscaled items as well as subscaled items, so this is also included in the analyses reported.

TABLE 2

Subscales of the Student Outcomes Checklist: Student Perceptions

Scale	Representative items (Students responded on a 5 point scale from "Applies a lot" to "Does not apply")	No. of items	Cronbach's alpha	F value (61, 633)
Mature inquiry	41. I learned that there is no single right answer to some scientific questions. 48. I learned how to work with other people.	9	.82	1.42
Science processes	15. I am better able to describe my observations. 19. I learned how to analyse data and tell what conclusions are justified by the data.	6	.79	2.00
Individual study skills	32. I learned to study on my own. 37. I learned to use a library to find information.	7	.83	1.81
Environmental awareness	16. I now realize that living things depend on each other. 33. I am more aware of some important problems that face mankind.	8	.81	1.72
Practical skills	8. I learned to use a microscope. 40. I learned to use lab. apparatus to get accurate results.	4	.57	1.87
Interest	44. I now read newspaper and magazine articles about biology. 55. I notice things around me that I never noticed before.	5	.73	1.44
Biological knowledge	6. I now have some understanding of the biologist's view of the world. 7. I learned how the different parts of an animal or plant work together.	8	.72	1.59
Total SOC	Total of subscales plus 13 unscaled items.	60	.94	1.96

Teacher-Pupil Relationships Scale (TPRS)

This instrument has evolved from one developed by Leeds (Leeds and Cook, 1947) for validating the *Minnesota Teacher Attitude Inventory.* It was used to get principal's ratings of teachers. Howe (1964) modified the instrument and later used it to get high school student's perceptions of their teachers. I then simplified the wording of Howe's instrument, added an extra item, and used it in U.S. Biology classes (Best, 1971).

When I was preparing to send out instruments to collect data at the end of 1974, I rang the teachers involved in the project and asked whether they would like this included. Most said "yes" and for these classes the TPRS (entitled *Student Inventory*) was attached to the answer sheets for both BAC:SP and SOC:SP.

The instrument consists of six items, each with five descriptions of a biology teacher – students are asked to tick the one most like their teacher. A high score indicates that the students like the teacher and see him or her as understanding and concerned with students as people. A low score indicates that students dislike the teacher, consider he or she has unrealistic expectations of them and is mainly concerned with subject matter.

The method of collecting the data meant that I had two sets of TPRS data, one from those students who had also answered BAC:SP and one from those students who had also answered SOC:SP. This meant that for classes where TPRS has been administered with both BAC:SP and SOC:SP, I had two estimates of the class mean on TPRS. The correlation between these two estimates (.87) convinced me that the instrument did measure something about the way the class sees the teacher, not just something about the way different individuals see the teacher. Results of the reliability estimates for TPRS are shown in Table 3. I will send copies of the instrument to anyone who wants one.

TABLE 3

Reliability estimates for Teacher-Pupil Relationships Scale

Sample	Cronbach's alpha	F value	Correlation of class means where classes contributed to both samples (49 classes)
Students answering BAC:SP	.79	6.1 (48, 477)	.87
Students answering SOC:SP	.79	7.0 (51,499)	

Analyses of Class Means

One would expect students' perceptions of the teacher, classroom activities and gains from doing biology to be interrelated. Table 4 shows correlations of class means for TPRS with subscale scores on the Biology Activity Report. Table 5 shows correlations of class means on TPRS and BAC:SP subscales with Student Outcomes Checklist subscale scores and total scores. Table 6 shows multiple correlations with SOC:SP subscale means as dependent variables and TPRS and Student initiative means as independent variables.

Discussion

The picture that emerges from these correlations seems to make sense. The students' perception of the teacher's personal concern is positively related to all outcomes except in-

TABLE 4

Correlations of Teacher-Pupil Relationships with Biology Activity Scales

BAC:SP Scale	With TPRS	BAC:SP Scale	With TPRS
Investigation	.32*	"Don't ask what it means, just get on with it."	-.43*
Student initiative	.44*	"Teacher knows best."	-.05
Demand for memorization	-.07	Free discussion	.37*

*$p < .01$ df 47

TABLE 5

Correlations of Activity Scales and Teacher-Pupil Relationships Scale with Outcome Scales

SOC:SP Scales BAC:SP Scales and TPRS	Mature inquiry	Science processes	Study skills	Env't awareness	Prac. skills	Interest	Knowledge	Total SOC
Investigation	.28*	.34**	.12	.31*	.19	.26*	.22	.32*
Student initiative	.52**	.49**	.34**	.35**	.46**	.44**	.33**	.51**
Demand for memorization	.27*	.18	.01	.21	.08	.14	.12	.22
"Don't ask what it means, just get on with it"	-.23	-.23	.02	-.07	-.22	-.25	-.24	-.19
"Teacher knows best"	.14	-.02	-.06	.03	.01	-.11	.13	.02
Free discussion	.21	.11	-.06	.14	.19	.11	.21	.12
Teacher-Pupil Relationships	.27*	.43**	.22	.38**	.18	.41**	.45**	.38**

*$p < .05$ df 56 for BAC:SP scales with SOC:SP scales.
**$p < .01$ df 51 for TPRS with SOC:SP scales.

dividual study skills and practical skills – both of which seem more related to individual work than to class work. Higher scores on the Investigation scale are positively related to outcomes having thinking process or attitudinal components, but not to individual study skills, practical skills or knowledge outcomes. On the other hand, higher scores on the student initiative scale are positively related to all outcomes including those that are related to individual work. The picture here is generally similar to the one I found in U.S. classes (Best, 1971) although the instrumentation was sufficiently different to make exact comparisons impossible.

One difference between the picture presented here and that seen in the U.S. data is that here there is a relationship between the students' perceptions of the teacher and their perceptions of classroom activities. (The instrument I used to measure classroom activities in

TABLE 6

Multiple correlations of Outcome Scales with Student initiative and TPRS as independent variables

SOC scale	Multiple R with TPRS and Student initiative	SOC scale	Multiple R with TPRS and Student initiative
Mature inquiry	.52**	Practical skills	.46**
Science processes	.55**	Interest	.51**
Study skills	.35*	Knowledge	.47**
Environmental awareness	.43**	Total SOC	.54**

*$p < .05$ **$p < .01$ df 2,48

the U.S. study was a one-shot report of a laboratory activity. The BAC:SP is more generalized.) The two activity variables that are not related to TPRS are Demand for memorization and "Teacher knows best". Neither of these variables seems to be related to anything else. (They are not correlated with other BAC:SP subscales either.) In the U.S. study a variable called "Teacher evaluation", analogous to "Teacher knows best", showed a similar lack of relationship to perceived outcomes.

Implications

It seems worth drawing the *Teacher-Pupil Relationships Scale* to the attention of researchers. It appears to be reliable both as a measure of individual students' perceptions of the teacher and as a measure of class perceptions. And it appears to measure one aspect of the classroom that is related to other things that are happening in it. It could perhaps be used to control for personal teacher effects in studies that are investigating teaching methodology. One danger in the use of the instrument is that teachers may find it threatening.

It also seems worthwhile to ask whether this study has anything to say to teachers. One problem with the TPRS is that it doesn't really give teachers information that they can do much about. On the other hand, teachers are certainly in a position to change the class score on the Student initiative scale of the Biology Activities Report if they wish to. (There is some suggestion that some teachers may have done this between 1973 and 1974). This study does not establish a causal relationship between higher scores on the Student initiative scale and higher scores on the outcome scales, but I believe it gives sufficient support to the value of encouraging student initiative to suggest to teachers that they try altering classroom activities in this direction. (All the student initiative items are shown in the Appendix).

A note of caution in conclusion. The relationships between biology activities and outcomes reported here are not apparent in the 1973 data. My best guess is that the within class sampling was at fault – in 1973 only one-third of the class answered each of BAC:SP and SOC:SP, resulting in very small samples from some classes. However another possibility is that the relationships cannot be generalized beyond the sample of classes that happened to be involved in 1974. The similarity to the relationships I found in the U.S. – and my own biases – suggest that the relationships are more general. The reader should make up his or her own mind – and perhaps classroom teachers may wish to see whether the relationships hold in their own classes.

References

BEST, E.D. The relationships among some components of student-centred inquiry methods and student's perceptions of the outcomes of biology teaching. *Research 1971*, Australian Science Education Research Association, 1971, 47-57.

BEST, E.D. Student perceptions of activities and outcomes in South Australian matriculation biology classes. *Research in Science Education*, 1974, **4**, 151-159.

HOWE, R.W. The relationship of learning outcomes to selected teacher factors and teaching methods in tenth grade biology classes in Oregon. Ann Arbor, Michigan: University Microfilms, 1964.

LEEDS, C.H. & COOK, W.W. The construction and differential value of a scale for determining teacher-pupil attitudes. *Journal of Experimental Education*, 1947, **16**, 149-159.

For data analyses I used the *Statistical Package for the Social Sciences*, and a merge program written by J. Saunders of the Research Section.

Appendix

Student initiative Subscale of the Biology Activity Report

Item	Percentage of students answering "yes"	
	1973	1974
28. We are encouraged to use the laboratory to investigate problems that come up in class.	27	42
32. We sometimes design our own experiments to seek answers to questions that puzzle us.	19	27
36. We never have the chance to try our own ways of doing the practical work.	55	51
41. We sometimes do some additional experimenting of our own when we have finished the set practical exercise.	18	25
47. All students in the class do the same exercises and the same problems.	72	73
49. We often work in groups for activities other than practical work.	49	51
50. We often have discussions in which students talk more than the teacher.	46	46
52. We are sometimes able to choose which problems we will discuss.	55	53
59. Students sometimes give short lectures or reports to the whole class.	19	24

Research in Science Education, 1977, 7, 113-121.

TEACHER CHARACTERISTICS AND PUPIL OUTCOMES IN SECONDARY SCIENCE CLASSROOMS

Jeffrey R. Northfield and Barry J. Fraser

Introduction

During 1974 the authors each initiated a different study related to the evaluation of the Australian Science Education Project (ASEP) at the seventh grade level in Victoria. An important component of one of the studies (Northfield, 1975) was a questionnaire survey involving science teachers, while a major orientation of the other study (Fraser, 1975, 1976a) was an investigation of relative changes experienced by ASEP and non-ASEP classes in cognitive and affective outcomes. Furthermore, in order to avoid the confusion associated with schools receiving requests to participate in two separate studies related to ASEP in the same year, a single letter was sent to each school requesting co-operation in either or both studies. This strategy led to overlap in the samples of schools involved in the two studies. As a result, there was a group of 31 teachers who had participated in the teacher questionnaire survey and whose students had completed the tests of cognitive and affective outcomes. The main purpose of the present paper is to report the use of data from this sample of 31 teachers and their classes in examining the relationship between certain teacher characteristics which were measured in the questionnaire survey and changes in pupil cognitive and affective outcomes.

Past research

Several comprehensive reviews of research (e.g. Dunkin & Biddle, 1974; Medley & Mitzel, 1963; Rosenshine & Furst, 1971) have generally revealed an absence of consistent findings of relationships between particular teacher characteristics and pupil learning outcomes. Nevertheless, whilst research has not led to the identification of teacher characteristics which consistently relate to student outcomes across a broad range of school contexts and types of students, numerous recent studies of science classrooms reviewed by Power (1977) have revealed some interesting relationships between teacher characteristics and student outcomes in specific instances.

It should be pointed out, however, that a large proportion of past studies into the relationship of teacher characteristics to learning outcomes has been conducted in classrooms following materials which are quite different from ASEP materials. Moreover, Tisher and Power (1976) have provided evidence that teachers do present science in ASEP classrooms differently from in conventional classrooms. It is possible, then, that the nature of ASEP materials could allow teachers more freedom to present science in their own preferred ways. Consequently, it would be interesting to explore the relationship between teacher characteristics and learning outcomes in classrooms following ASEP materials.

Moreover, with the current availability of a wide variety of science curriculum materials and the increasing freedom for Australian teachers to make choices among those alternatives, it is possible that relationships between teacher characteristics and pupil outcomes could vary with the instructional materials used. It is intuitively plausible that different teachers would vary in their preference for different instructional materials, and that optimal combinations of certain instructional materials with teachers of particular characteristics could lead to the promotion of more favourable pupil outcomes than other combinations. In the present study, then, a major focus was seen to be the investigation of possible interactions between an instructional treatment variable (involving the use of either ASEP or alternative materials) and teacher characteristics in the prediction of pupil outcomes.

Choice of teacher characteristics

In past research, teacher variables have often been chosen for study without a sound justification for expecting a relationship to exist between those teacher variables and pupil outcomes. A justification is given below for the inclusion in the present research of each of the following four teacher variables: *sex, experience, attitude to pupil-centredness* and *attitude to structure.*

Teacher sex was included in the present investigation because various studies have indicated that male and female teachers differ in their classroom behaviour and attitudes to teaching. For example, in their review of Ryan's study of teachers, Getzels and Jackson (1963, p.568) reported that "differences between the sexes were fairly general and pronounced with women tending to attain significantly higher scores on the seven scales measuring friendly, responsible, stimulating classroom behavior, favorable attitudes towards pupils, democratic classroom practices, permissive educational viewpoints, and verbal understanding".

The second teacher variable included in the research is the number of years of teaching experience. It could be expected that more experienced teachers might have developed more teaching skills than less experienced ones, although much of this experience would probably have been gained in conventional classrooms. On the other hand, less experienced (and therefore usually younger) teachers would be more likely to have experienced some innovative science curricula in their own years as school students or to have received training in the use of innovative programs during their teacher education courses. In view of these arguments, it is possible that teacher experience, or the interaction of the instructional variable with teacher experience, might be related to pupil outcomes. Nevertheless, it is interesting to note that a study by Rothman, Welch and Walberg (1969) showed that teacher experience was not significantly related to the achievement of numerous cognitive and affective aims of a senior physics course.

The last two teacher characteristics included in the present investigation are attitude to pupil-centredness and attitude to structure. These two attitudes were measured, respectively, by the 10-item and five-item factor analytic scales described by Fraser and Northfield (1976). Whereas high scores on the attitude to pupil-centredness scale would involve agreement with statements that teachers should "have pupils spend most of their time working in small groups" or "allow pupils the opportunity to select some of their activities during the year", high scores on the attitude to structure scale would involve agreement with statements that a teacher should "provide a year's course which forms part of a clearly planned Form I to IV program of science" or "present ideas from each of the major scientific disciplines".

As mentioned previously, it is reasonable to anticipate that certain teacher characteristics and instructional materials could interact to predict pupil learning outcomes. Since teachers' attitudes to specific aspects of science teaching would be an important determinant of the way teachers actually implement a curriculum in the classroom, it could be anticipated that teacher attitude to pupil-centredness or structure could interact with the present instructional variable (ASEP/non-ASEP) in predicting some student learning outcomes. This expectation is further supported by the theoretical position of Boulding (1956). Boulding has introduced the concept "image", which refers to the cognitive and affective map of the world possessed by every person, which provides the means by which an individual interprets the world. House (1973) has applied Boulding's concept of image to teachers in the following way:

> Since all information is filtered through the value system that is entwined with the image, there are no 'facts' – only messages that are allowed to filter through. The image is impermeable to change. (House, 1973, p.257)

It is consistent with this theoretical position that the teacher's image of science teaching would be stable and influential in affecting science teaching behaviour. The present variables of teacher attitude to pupil-centredness and structure could be considered components of the teacher's image of science teaching and, as such, determinants of teaching behaviour and, indirectly, pupil outcomes.

Learning outcome measures

The seven learning outcome measures employed in the present analyses, together with α reliability estimates, are shown in Table 1. These seven measures, which have been discussed in detail in Fraser (1974), embrace the 17 individual scales used in previous analyses (Fraser, 1975, 1976b). The first measure, namely Test of Enquiry Skills (TOES) Part A, includes the two skills of library usage, and use of a book's index and table of contents. TOES Part B includes the four skills of reading scales, finding averages, percentages and proportions interpreting charts and tables, and interpreting graphs. TOES Part C measures the three abilities of comprehension of science reading material, design of experimental procedures, and forming conclusions and generalizations from experimental data. The fourth measure in Table 1, namely Test On Understanding Science (TOUS), covers understanding of philosophical aspects of science, of historical-social aspects of science, and that scientists are normal people. The fifth measure covers three attitudinal outcomes, namely attitude to the social implications of science, enjoyment of science lessons and interest in science outside lessons. The last two instruments in Table 1, which measure attitude towards using enquiry methods to find out information and adoption of certain scientific attitudes like curiosity, respectively, are identical to the scales used in previously reported analyses. Although separate analyses were performed initially for the 17 individual scales, it was found that data could be reported only for the seven combined measures in Table 1 with a considerable gain in economy and with negligible loss in information yield.

Design of the study

The sample consisted of 31 teachers and their 31 seventh grade science classes (780 students), each in a different coeducational high school in the Melbourne metropolitan area. Although schools were not randomly chosen for a variety of reasons, they were still spread representatively over the range of geographic and socio-economic areas around Melbourne. Of the 31 classes, 15 classes formed the ASEP group while the remaining 16 classes formed

TABLE 1

Percentage of Variance in each of seven Learning Outcome Posttests explained by Full Regression Model and by Pretest, and Percentage of Variance Uniquely explained by four Teacher Variables and four Instruction x teacher Interactions

Learning Outcome	α Reliability	R^2(%) for Full Model	R^2(%) for Pre-test	R^2(%) for Teacher Variables				R^2(%) for Instruction x teacher Interactions			
				Sex	Experience	Pupil centredness	Att to structure	Instr x sex	Instr x Experience	Instr x pupil centredness	Instr x attitude to structure
TOES Part A: Reference Materials	0.65	62.2	52.4	0.0	0.1	0.4(.1)	0.2	0.7*	0.3	0.3	0.4
TOES Part B: Interpreting & Processing Information	0.91	73.1	67.9	0.0	0.0	0.3(.1)	0.3(.1)	0.2	0.5*	0.4(.1)	0.1
TOES Part C: Critical Thinking in Science	0.85	66.0	57.7	0.0	0.4(.1)	L0.6*	0.0	0.0	0.0	0.0	0.0
TOUS: Test on Understanding Science	0.78	53.0	42.0	0.5(.1)	0.1	0.0	0.8(.1)	0.2	0.8(.1)	0.0	0.1
Attitude to Science	0.89	35.2	26.6	F1.9**	0.3	M1.7*	0.1	0.1	0.2	0.4	0.0
Attitude to Enquiry	0.72	25.2	14.4	0.1	0.1	0.3	1.0(.1)	0.1	0.7	1.8*	1.2(.1)
Adoption of Scientific Attitudes	0.50	43.6	30.0	0.6(.1)	0.1	0.0	0.0	0.2	0.0	0.4	2.7**

(.1) $p < .1$, *$p < .05$, **$p < .01$

Sample size was 231 subgroups (31 schools)

Full regression model contains 24 terms: Pretest, instruction (ASEP/control), SES, IQ, sex, 11 interactions, 4 teacher variables and 4 instruction x teacher interactions

F Female teachers' classes experienced more favourable changes than male teachers' classes

L Less pupil-centred teachers' classes experienced more favourable changes than more pupil-centred teachers' classes

M More pupil-centred teachers' classes experienced more favourable changes than less pupil-centred teachers' classes the control group. Teachers in the ASEP group used ASEP materials exclusively during the time of the study, but the choice, sequencing and timing of units were left to each teacher's discretion. The control treatment for the present study was also broadly defined in terms of the use of any science instructional materials, with the exception of ASEP materials.

The 780 students in the sample each answered the battery of tests in Table 1 as pretests in March, 1974 and again as post-tests in October. The 31 teachers in the sample answered a questionnaire in February or March, 1974, and this provided data on the four teacher characteristics of sex, experience, attitude to pupil-centredness and attitude to structure. As well, data on student socio-economic status (SES), general ability (IQ) and sex were gathered in August. While SES was measured using Congalton's (1969) classification of fathers' occupations, IQ was measured with a modified version of the Otis test (Mackay and Spicer, 1975).

For reasons given in Fraser (1975), the actual unit of statistical analysis employed in the present study was not the individual pupil but the subgroup within the class formed by grouping students according to similarities in SES, IQ and sex. This sampling strategy involves the division of the pupils within each class according to their SES, IQ and sex. (In fact, the eight groups were high SES – high IQ – boys, high SES – high IQ – girls, . . . , low SES – low IQ – girls.) It was found that, either because of student absences during testing or because of the distribution of SES, IQ and sex in some classes, the total sample yielded 231 subgroups which is only 17 less than the maximum number possible if each of the 31 classes provided all eight subgroups.

Data analysis

A multiple regression analysis was performed separately for each learning outcome post-test as criterion. Each full regression model consisted of the following 24 items: the corresponding pretest, the instructional variable (ASEP/control), the three pupil aptitudinal variables (SES, IQ and pupil sex), the four teacher characteristics (teacher sex, experience, attitude to pupil-centredness and attitude to structure), the 11 interactions possible between the four variables of instruction, SES, IQ and pupil sex, and the four instruction-teacher interactions between instruction and one of the four teacher characteristics. The statistical power of the analyses was maximized by maintaining all quantitative variables as such (instead of reducing them to categories), both when estimating main effects and when forming interactions by taking products of predictor variables, as recommended by Cohen and Cohen (1975). The variance-partitioning algorithm employed was Overall and Spiegel's Method 2 (see Overall & Spiegel, 1969; Kaufman & Sweet, 1974; McArthur, West & Fraser, 1975). In Method 2, the amount of variance associated with each main effect is estimated after removal of the variance attributable to all other main effects, but not interactions; then the variance associated with each two-way interaction is estimated after removing the variance attributable to all main effects and all other two-way interactions, but not to higher order interactions; etc.

The second column of figures in Table 1 shows that the amount of variance in a learning outcome post-test accounted for by the full 24-term model ranged from 25.2 to 73.1 per cent. The third column of figures in the table indicates that the amount of variance in a learning outcome post-test accounted for by the pretest (i.e. the square of the simple pretest-post-test correlation) was quite large and ranged from 14.4 to 67.9 per cent, thus indicating that a relatively small amount of variance remained to be partitioned amongst all other predictor variables. The remaining columns in Table 1 show the percentages of variance uniquely associated with each teacher variable and with each instruction-teacher interaction, together with the results of significance tests, when Method 2 was employed. For example, the fourth column contains the percentage of variance in each post-test uniquely accounted for by teacher sex, after the variance due to pretest, instruction, pupil SES, IQ and sex, and the other three teacher variables had been removed. The figures in the last column of Table 1 represent the percentage of variance in each post-test uniquely associated with the instruction-attitude to structure interaction, after the removal of variance attributable to pretest, instruction, SES, IQ and pupil sex, the four teacher characteristics, and all other two-way interactions contained in the full model.

The results in Table 1 for the four teacher main effects indicate that a teacher variable was a significant independent predictor of changes in a learning outcome in three out of 28 cases at the 0.05 level of significance (about twice that expected by chance), and in 11 out of 28 cases at the 0.1 level of significance (about four times that expected by chance). Results for the four instruction-teacher interactions show that an interaction term was a significant independent predictor of changes in a learning outcome in four out of 28 cases at the 0.05 level of confidence (about three times that expected by chance), and in seven out of 28 cases at the 0.1 level of confidence (about three times that expected by chance). Although only relationships significant at the 0.05 level of confidence are interpreted below, those significant at the 0.1 level have been recorded in Table 1 to illustrate that both teacher characteristics and instruction-teacher interactions tended to be relatively good predictors of changes in pupil learning outcomes even after the effect of instructional and pupil aptitudinal variables had been removed. Moreover, the reasonable proportion of significant instruction-teacher interactions in the present study can be compared with the paucity of instruction-aptitude interactions revealed in a number of research reviews (Bracht, 1970; Cronbach & Snow, 1969).

The three cases in which a teacher characteristic was found to be a significant independent predictor were for teacher sex and Attitude to Science, attitude to pupil-centredness and TOES Part C, and attitude to pupil-centredness and Attitude to Science. The interpretations of these three cases, in which the hypothesis of no relation between a teacher variable and changes in a learning outcome was rejected, are given below. Declines in Attitude to Science scores were smaller (i.e. more favourable) in female teachers' classes than in male teachers' classes. Improvement on TOES Part C (Critical Thinking in Science) was greater in classes of less pupil-centred teachers than in classes of more pupil-centred teachers. Declines in Attitude to Science were smaller (i.e. more favourable) in classes of more pupil-centred teachers than in classes of less pupil-centred teachers.

The four cases in Table 1 for which an instruction-teacher interaction was a significant independent predictor of changes in a learning outcome were instruction-teacher sex and TOES Part A, instruction-experience and TOES Part B, instruction-attitude to pupil-centredness and Attitude to Enquiry, and instruction-attitude to structure and Adoption of Scientific Attitudes. The interpretation of these four interactions is depicted graphically in Figure 1. Student improvement on TOES Part A (Reference Materials) was independent of teacher sex in the ASEP group, but control pupils with female teachers experienced a greater improvement than control pupils with male teachers. Student improvement on TOES Part B (Interpreting & Processing Information) decreased with teacher experience in the ASEP group, but increased with teacher experience in the control group. Student improvement on the Attitude to Enquiry scale increased with the pupil-centredness of teachers in the ASEP

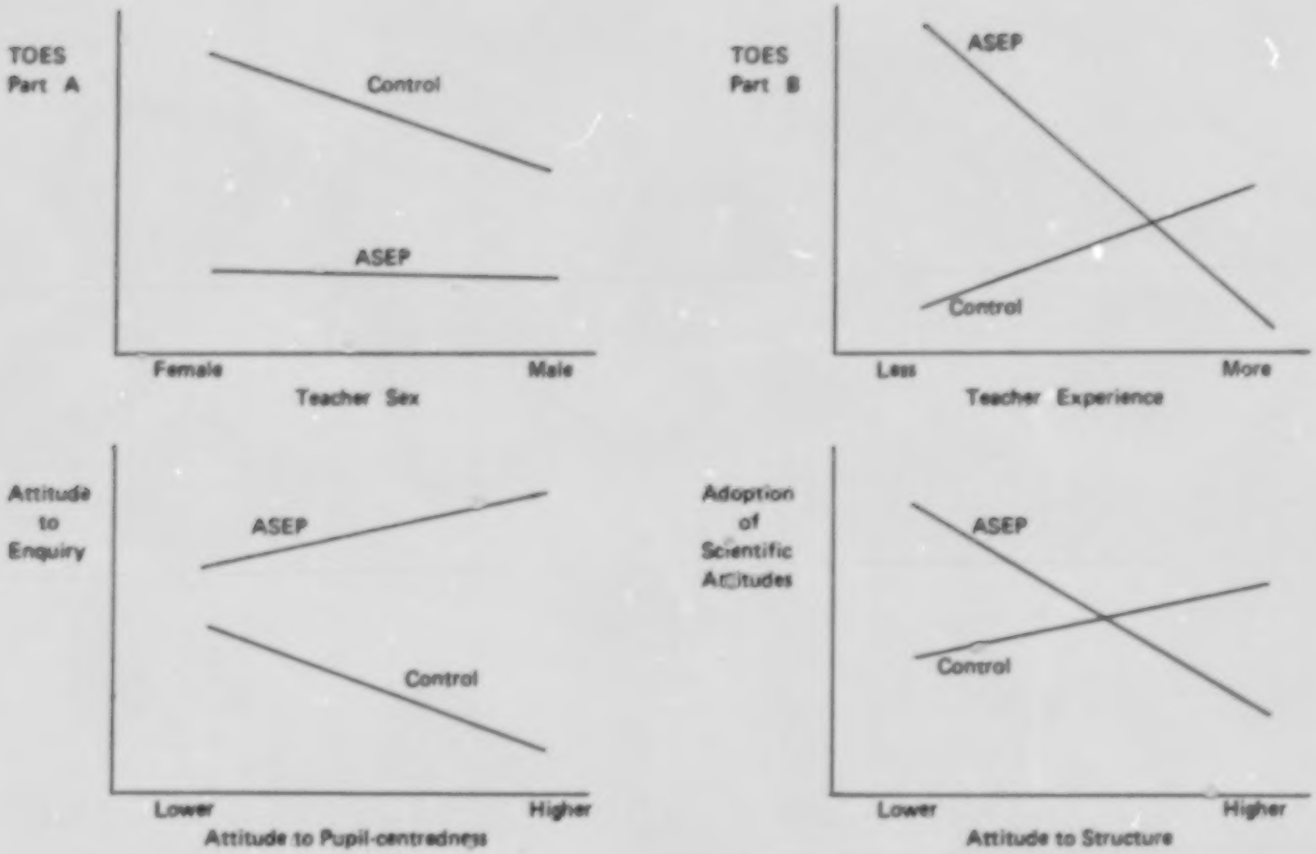


FIGURE 1

Instruction-teacher Interactions

group, but decreased with the pupil-centredness of teachers in the control group. Student improvement on the Adoption of Scientific Attitudes scale decreased with teacher attitude to structure in the ASEP group, but increased with teacher attitude to structure in the control group.

Conclusion

The present research employed data collected in two separate studies related to ASEP to investigate relationships of changes in seven pupils learning outcomes with four teacher characteristics and four instruction-teacher interactions. A significant relationship ($p < .05$) emerged between changes in a learning outcome and a teacher characteristic in the following three cases: changes in critical thinking in science were more favourable in classes with less pupil-centredness teachers, and changes in attitude to science were more favourable in classes with female teachers and those with more pupil-centredness teachers. A significant relationship also emerged between changes in a learning outcome and the interaction of the instructional variable (ASEP/control) with a teacher characteristic in four cases, and these were interpreted in the paper.

In addition to the specific findings outlined above, two more general tentative conclusions can be advanced. First, the use of common samples in separate studies appears useful for allowing the investigation of worthwhile questions which go beyond the bounds of the individual studies. Second, in view of the number of interesting relationships found between changes in learning outcomes and teacher and instruction-teacher interaction variables, it could be useful to replicate the present study.

References

BOULDING, K.E. *The image*. Ann Arbor: University of Michigan, 1956.

BRACHT, G.H. Experimental factors related to aptitude-treatment interactions. *Review of Educational Research*, 1970, 40, 627-647.

DUNKIN, M., & BIDDLE, B. *The study of teaching*. New York: Holt, Rinehart & Winston, 1974.

COHEN, J., & COHEN, P. *Applied multiple regression/correlation analysis for the behavioral sciences*. New York: John Wiley, 1975.

CONGALTON, A.A. *Status and prestige in Australia*. Melbourne: Cheshire, 1969.

CRONBACH, L.J., & SNOW, R.E. Individual differences in learning ability as a function of instructional variables. Unpublished report, Stanford University, March 1969.

FRASER, B.J. Selecting evaluation instruments. *Research in Science Education*, 1974, 4, 99-111.

FRASER, B.J. The impact of ASEP on pupil learning and classroom climate. *Research in Science Education*, 1975, 5, 1-12.

FRASER, B.J. An evaluation of ASEP involving learning outcome, aptitudinal and environmental variables. Unpublished Ph.D. thesis, Monash University, 1976a.

FRASER, B.J. Classroom climate as predictor and criterion in science education research. *Research in Science Education*, 1976b, 6, 109-120.

FRASER, B.J. & NORTHFIELD, J.R. A study of ASEP during its first year of availability. Report to the Curriculum Development Centre, October, 1976.

GETZELS, J.W., & JACKSON, P.W. The teacher's personality and characteristics. In N.L. Gage (Ed.), *Handbook of research on teaching.* Chicago: Rand McNally, 1963.

HOUSE, E.R. The relevance of evaluation. In R.M. Rippey (Ed.), *Studies in Transactional Evaluation.* Berkeley: McCutchan, 1973.

KAUFMAN, D., & SWEET, R. Contrast coding in least squares regression analysis. *American Educational Research Journal*, 1974, 11, 359-377.

MACKAY, L.D., & SPICER, B. Educational turbulence among Australian servicemen's children. Unpublished ETASC project report, Volume 1, Monash University, April, 1975.

McARTHUR, J.T., WEST, L.H.T., & FRASER, B.J. Some issues in the simulation of ANOVA by multiple regression techniques. *Research in Science Education*, 1975, 5, 199-207.

MEDLEY, D.M., & MITZEL, H.E. Measuring classroom behavior by systematic observation. In N.L. Gage (Ed.), *Handbook of research on teaching.* Chicago: Rand McNally, 1963.

NORTHFIELD, J.R. Changing perspectives of science teacher education that may be associated with curriculum development. *Research in Science Education*, 1975, 5, 23-32.

OVERALL, J.E., & SPIEGEL, D.K. Concerning least squares analysis of experimental data. *Psychological Bulletin*, 1969, 72, 311-322.

POWER, C.N. A critical review of science classroom interaction studies. *Studies in Science Education*, 1977, 4, 1-30.

ROSENSHINE, B., & FURST, N. Research on teacher-performance criteria. In B.O. Smith (Ed.), *Research on teacher education.* New Jersey: Prentice Hall, 1971.

ROTHMAN, A.I., WELCH, W.W., & WALBERG, H.J. Physics teacher characteristics and student learning. *Journal of Research in Science Teaching*, 1969, 6, 59-63.

TISHER, R.P., & POWER, C.N. Variations between ASEP and conventional learning environments. *Australian Science Teachers Journal*, 1976, 22(3), 35-39.

Research in Science Education, 1977, 7, 123-129.

SEX DIFFERENCES IN STUDENT PREFERENCES FOR, AND PERCEPTIONS OF, LEARNING OUTCOMES AND CLASSROOM ACTIVITIES IN YEAR 11 BIOLOGY

Janet V. Keightley

Introduction

The effects of student sex differences on science education have been investigated from a number of aspects. In a review of the literature by Gardner (1974) it is noted that there is much evidence to indicate that boys prefer to study physical science topics while girls tend to prefer topics in the biological and social science areas. This study investigates the effect of sex differences of those students who choose to study a biological science subject.

Procedures

Development of checklist

1. *Student preferences:* The checklist used by Keightley and Best (1975) was modified to expand the total number of items to forty-four. The examples in that checklist were made into individual items. Items concerning independent learning and environmental awareness were added. A second part of the checklist included ten possible activities in a biology classroom and one item asking how interesting the student thought the biology course would be.

   The data discussed in this paper was collected by the original checklist in 1975 and the modified one in 1976.

2. *Student perceptions:* The wording of each item in the student preference checklists for that year was changed so that the item could be viewed as an outcome which may have been achieved to a varying degree in each biology class. The classroom activities were also included in the perception checklist with similar wording changes and an item asking how interesting the student had found biology to be.

3. *Student interest:* In both checklists one item relating to student interest was included. The item asked the students to apply a five point rating scale of "Very interesting", "Fairly interesting", "Alright", "Rather boring" and "Very boring" to their expectations of how interesting they hoped the biology course to be, and to their experience of the course.

Administration

At the beginning of the academic year, all biology teachers in South Australia were invited to participate in the project. Those who returned completed student questionnaires at the beginning of the year were sent perception questionnaires at the end of the year.

Responses

Students were requested to respond to each item on a four point rating scale. The scale for the preferred learning outcomes was "Very much like to learn", "Like to learn", "Don't care" and "Rather not learn". This scale was changed to "Applies a lot", "Applies moderately", "Applies a little", and "Does not apply" for the perceptions checklist at the end of the year. The four point rating scale used for the classroom activities was a time scale of "Never", "Less than once a month", "Once or several times a month", and "Once a week or more" and students indicated their preferred frequency at the beginning of the year and their perceived frequency at the end of the year. The item asking how interesting the students expected the course to be and how interesting they found the course involved a five point rating scale of "Very interesting", "Fairly interesting", "Alright", "Rather boring" and "Very boring".

The distribution of responses collected is shown in Table 1.

TABLE 1

Numbers of student responses

Student preferences

	Students	Total Classes	All Females Classes	All Males Classes
1975	2 620	121	12	7
1976	1 729	86	8	2

Student perceptions

	Students	Total Classes	All Females Classes	All Males Classes
1975	1 170	79	10	3
1976	1 151	65	6	3

Results

For each item the mean for the male population and the mean for the female population were computed and a t-test of significance applied to the difference between the means. For student preferences, the results are shown in Tables 2 and 3. An examination of the data on student perceptions revealed that there were no male means significantly greater than the female means. Table 4 show those items for which the female means were significantly greater than male means.

Sex differences in Student preferences for classroom activities are shown in Table 5. Male and female perceptions of classroom activities differed to a much smaller extent. In

TABLE 2

Student preferences for learning

Male means significantly greater than females means

1976

To learn to dissect animals (-0.*22).
To learn to use laboratory apparatus (-0.25).

1975 and 1976

To learn to think in a scientific way (-0.16 and -0.25).
To learn practical skills (-0.18 and -0.13).

1975 Mean Scale between 2.40 and 3.40; 0.16 Mean Difference significant at .001 level.
1976: Mean Scale between 2.46 and 3.40; 0.11 Mean Difference significant at .001 level.

*Mean Difference — a positive value indicates a higher female mean.
— a negative value indicates a higher male mean.

TABLE 3

Student preferences for learning outcomes

Female means significantly greater than male means

1976

To learn how animals, plants and man affect each other (+0.*11).
To learn the names of different parts of plants and animals (+0.19).
To learn details about how different parts of plants and animals work (0.13).
To gain a greater interest in living things (+0.22).
To gain a greater appreciation of and respect for nature (0.25).
To learn to explain my ideas to others (+0.21).
To learn to work together with others on a problem (+0.18).
To learn to see other people's points of view (+0.22).
To develop an awareness of the environment (+0.15).
To become more aware of the important problems facing mankind (+0.13).

1975 and 1976

To develop interests and attitudes (+0.17 and +0.15).
To gain biological knowledge with everyday application (+0.19 and 0.16).
To gain knowledge and skills useful for a job (+0.29 and +0.17).
To learn to co-operate (+0.21 and +0.14).

Not significant 1975, significant 1976

To gain a basis for further learning (+0.10 and +0.19).
To gain an understanding of the nature of science (-0.16 and +0.13).
To learn to communicate (+0.15 and +0.23).

1975: Mean Scale between 2.40 and 3.40; 0.16 Mean Difference significant at .001 level.
1976: Mean Scale between 2.46 and 3.40; 0.11 Mean Difference significant at .001 level.

*Mean Difference — a positive value indicates a higher female mean.
— a negative value indicates a higher male mean.

TABLE 4

Student perceptions of learning outcomes

Female means significantly greater than male means

1975 and 1976

To learn how animals, plants and man survive (+0.20* and +0.16).
To gain a knowledge of detailed facts about parts and animals (+0.25 and +0.23).
To learn the names of different parts of plants and animals (+0.28 and +0.21).
To learn to organize my work (+0.22 and +0.23).
To develop interests and attitudes (+0.27 and +0.18).
To gain a greater interest in living things (+0.30 and +0.28).
To gain a greater appreciation of and respect for nature (+0.22 and +0.17).
To learn some things that might help me to choose foods for a healthy diet (+0.29 and +0.25).
To gain a basis for further learning (+0.21 and +0.27).
To gain a basis for matriculation biology (+0.26 and +0.22).
To gain knowledge and skills useful for a job (+0.37 and +0.36).
To gain an understanding of the nature of science (+0.17 and +0.17).
To learn how to write clearly (+0.22 and 0.32).
To learn to work together with others on a problem (+0.25 and +0.29).
To learn to see other people's points of view (+0.23 and +0.30).

Significant 1975, not significant 1976

To learn how science has affected human history (+0.21 and +0.12).

Not significant 1975, significant 1976

To learn to explain my ideas to others (+0.15 and +0.18).
To learn to co-operate (+0.14 and +0.28).

Data not available 1975, significant 1976

To develop an awareness of the environment (+0.25).

1975: Mean Scale between 1.95 and 3.48; 0.17 Mean Difference significant at .001 level.

1976: Mean Scale between 1.70 and 3.51; 0.17 Mean Difference significant at .001 level.

---

*Mean Difference – a positive value indicates a higher female mean.
– a negative value indicates a higher male mean.

TABLE 5

Student preferences for frequency of classroom activities

Male means significantly greater than female means

1975 and 1976

Films (-0.*29 and -0.24)
T.V. programmes (-0.33 and -0.27).
Colour slides and other A.V. aids (-0.21 and -0.14).
Field Work (-0.18 and -0.18).

Female means significantly greater than male means.

1975 and 1976

Teacher talking (+0.30 and 0.26).
Students working individually (+0.28 and +0.29).

No significant difference between means of males and females.

1975 and 1976

Visiting speakers (+0.01 and +0.06).
Small group discussions (other than practical) (+0.05 and +0.12).
Whole class discussion (+0.06 and +0.07).
Students reporting to class (-0.06 and -0.01).

1975: Mean Scale between 1.84 and 3.59; 0.18 Mean Difference significant at .001 level.

1976: Mean Scale between 1.87 and 3.54; 0.14 Mean Difference significant at .001 level.

*Mean Difference – a positive value indicates a higher female mean.
– a negative value indicates a higher male mean.

1975, the male mean was significantly greater than the female mean (-0.18 Mean Difference; Mean Scale 1.22-3.52) only for the item "T.V. Programmes" while in 1976 the only item where the sexes differed significantly in their perceptions of frequency of classroom activities was "Small group discussions" where the female mean was significantly greater than the male mean. (Mean Difference +0.26; Mean Scale 1.16-3.52).

The means for females on the interest item were significantly greater than those of the male students in both the expected and experienced interest. (Table 6).

TABLE 6

Student interest

Anticipated interest

	Female Mean	Male Mean	Mean Difference
1975	4.25	4.05	+0.20*
1976	4.19	3.92	+0.27*

Experienced interest

	Female Mean	Male Mean	Mean Difference
1975	3.90	3.50	+0.40*
1976	4.13	3.76	+0.37*

*Significant at .001 level.

Discussion of data

In a study by Walberg (1967) of physics students, girls were shown to score significantly higher than boys on three of five dimensions he identified. These factors were "Academic" which included reading, memorizing, writing reports and discussing science in class; "Nature Study" which included visiting zoos and gardens and feeding wild birds; and "Applied Life" which included practical interest in living things, scientists, plants and animals.

The results of this study are consistent with Walberg's findings. In Table 3, female preferences were significantly greater on items relating to academic pursuits (e.g. names of different parts of plants and animals, details of how different parts of plants and animals work), nature (gain a greater appreciation of and respect for nature and gain a greater interest in living things), as well as those with practical application (learn some things that might help me to choose foods for a healthy diet, knowledge and skills useful for a job.

In the same study Walberg found that boys scored significantly higher than girls on two factors called "Tinkering" which included practical and experimental items and "Cosmology" which included more theoretical and speculative aspects. In this study (Table 2) male means were significantly greater in items relating to practical activities (to dissect animals, use laboratory apparatus) and more theoretical aspects (to think in a scientific way).

These consistent findings in choice of learning outcomes have implications which should be considered in discussions on individualizing instruction. The significant differences in means for preferred classroom activities should also be considered. In this study (Table 5) males favour an audiovisual approach with field work while females preferred individual work with teacher lecturing.

The different preferred orientations of males and females is one aspect of individual differences which should be identified and seriously considered when student learning programmes are being developed. Also, when setting up a classroom learning environment, the preferred kinds activities should be considered.

When students were asked their perceptions of the degree to which each learning outcome applied to their course, for almost half the items the female means were significantly greater than the male means (Table 5). There were no male means significantly greater than female means. When students were asked to indicate their expected and experienced interest in the course, the female means in all cases were significantly higher than the male means. The female students show an overall more positive attitude towards the subject. In a study done in New Zealand by Fitt (1956) girls showed a greater liking for school than boys did and therefore it seems consistent that this overall enthusiasm would be reflected in their perceptions of learning outcomes achieved and the interest the course held for them.

The significant difference in the perceptions of frequency of two classroom activities between males and females (T.V. programmes and small group discussions) reveals the need for further investigation into biology classroom activities and the degree of participation and involvement by males and females in the class.

References

FITT, A.B. An experimental study of children's attitude to school in Auckland, N.Z. *British Journal of Educational Psychology*, 1956, 26(*1*), 25-30

GARDNER, P.L. Sex differences in achievement, attitudes and personality of science students: A review. *Research in Science Education*, 1974, 4, 231-258.

KEIGHTLEY, J.V. & BEST, E.D. Student preferences for year 11 biology classes in some South Australian schools. *Research in Science Education*, 1975, 5, 59-67.

WALBERG, H.J. Dimensions of scientific interest in boys and girls studying physics. *Science education*, 1967, 51, 111-116.

Research in Science Education, 1977, 7, 9-24.

LOGICAL CONNECTIVES IN SCIENCE : A SUMMARY OF THE FINDINGS

P. L. Gardner

> Hydrogen stands alone among the elements of the periodic system. Its single electron might make feasible its classification with the alkali metals but hydrogen does not lose its electron easily enough to warrant such classification. Similarly, hydrogen has at times been classified with the halogen elements because it needs only one more electron to satisfy the rare-gas electron arrangement of helium. In any case, trying to classify hydrogen as an alkali metal or as a halogen would only be "forcing" such a classification; for this reason, hydrogen will be considered unique among the elements, constituting a "family" in itself.
>
> Nachamkin (1968, p.1)

This passage, the opening paragraph of the first chapter of a chemistry text, is a fairly typical sample of scientific writing. The paragraph is composed of a number of related propositions. These propositions, and the concepts out of which they are constructed, are linked together to form a coherent argument.

Linkage in English is achieved in many different ways. For example, one form of linkage, known as hyponymy (class-inclusion), is demonstrated in Sentences 1 and 2 of the quoted passage: the 'alkali metals' of Sentence 2 are sub-class of the 'elements of the periodic system' mentioned in Sentence 1. A second form of linkage is brought about by the many function words (e.g. *'among', 'its', 'with', 'such'*) which tie concepts together within a sentence, or carry meaning from one sentence to another.

A third form of linkage is achieved by logical connectives. These are words and phrases which serve as links between sentences, or between a clause and either a phrase or another clause within a sentence. Examples in the quoted passage are *'but', 'similarly', 'because', 'in any case'*, and *'for this reason'*. It is this kind of linkage which is the subject of the present research. The Logical Connectives in Science Project has sought to identify those logical connectives which occur most frequently in scientific material, and to investigate secondary school students' difficulties with these terms. The method of compilation of the list of logical connectives, the types of test items employed to investigate difficulties, and some of the preliminary findings from the pilot stages of the project, have been described in two earlier papers (Gardner, 1975, 1976). During 1976, the final versions of the tests were prepared and administered, and the results analysed. A summary of this final phase of the project is given in the present paper.

Testing and sampling

The tests

The final set of tests consisted of 25 different tests, each containing 40 multiple-choice items. Amongst the thousand items in the item bank, half were gap-filling (GF) items, and nearly all the rest were sentence-completion (SC) items. In about half the items, the logical connective was tested in a scientific (S) context; in the remainder it was tested in an everyday (E) context. The following four items testing the work 'actually' are presented for the purposes of illustration:

1. attitudes of enquiry — the *habit* of critical thinking;
2. knowledge of the nature of situations in which the weight or accuracy of statements is to be logically determined;
3. mental skills in applying attitudes and knowledge.

It is the second aspect which lends support to the suggestion that different critical thinking tests are required for different subject areas, for it can be argued that many science situations are different from non-science ones, since they contain aspects of empirical enquiry not common to other areas of knowledge.

Development of the initial test

It was clear that, in order to test aspect (3) using a pencil and paper instrument, I could not expect to improve upon the well-tried approach of the WGCTA. Aspect (1) cannot genuinely be assessed in this manner, but I have proceeded on the assumption that a composite of aspects (2) and (3), which might be termed *cognitive critical thinking ability*, can be validly tested by providing statements referring to material characteristic of the area of study, in this case science, and requiring assessment of these statements by means of evidence and logic. Part of the evidence may be contained within the test material, but part may be required to be contributed by the respondent from his general knowledge, and understanding of the nature of the science subject. General knowledge is a requirement for some items in the WGCTA. The WGCTA consists of five subtests of different format designed to tap somewhat different aspects of the skills required. This subtest structure was retained in the science-based tests used in the work described in this paper.

Since chemistry students were the subject of the study, the first test constructed contained material dealing mainly with chemistry situations. This test was initially part of a battery of tests administered to groups of students, and teachers were not able to make available the double class period necessary for a test of the length of the WGCTA, which has 100 items. For this reason a test of 42 items was produced; it was intended to take 30-35 minutes and will be referred to as Form C.

Examples of WGCTA items, with corresponding chemistry-based items and items from a subsequent physics version (Form P) to be discussed below, are given in Table 1.

Experimental results

For a test in which most of the items are essentially of the true-false type, reliability can be a problem, and in early trials in 1975 this was the case. The test was developed for curriculum evaluation purposes, not for individual assessment, so the low reliability, which in any case could clearly become respectable if the number of items were increased, was not of major importance. However, it was decided for several reasons to produce an alternate form whose only major difference from the first was that most situations were related to physics. Both tests were administered, one week apart in early 1976, to a group of students and the results were used to estimate alternate-forms reliability. Internal-consistency and product-moment correlation data are shown in Table 2.

Form C was administered first, and some students did not finish the final subtest. If this is taken into account, the means are almost identical. This has subsequently remained

TABLE 1

Examples of test items

WGCTA, Form Ym	Recognition of Assumptions	*Stem:* "A wise man will save at least twelve dollars each week out of his earnings". *Proposed assumption:* (is this made in the above statement?): "No fools have sense enough to save twelve dollars a week".
Form P	Recognition of Assumptions	*Stem:* "The continued burning of fossil fuels will cause a catastrophic rise in the earth's temperature due to the insulating effect of increased concentration of atmospheric carbon dioxide". *Proposed assumption:* "Carbon dioxide produced by burning fossil fuels cannot be prevented from entering the atmosphere".
Form Ym	Interpretation	*Stem:* "A report of the U.S. Census during a certain year states that there were approximately 1,656,000 marriages and 264,000 divorces granted in the United States". *Conclusion:* (does this follow from the above?) "The divorce rate in the U.S. is much too high".
Form C	Interpretation	*Stem:* "A chemist, in trying to make two substances react together, used every likely method but to no avail. He gave up, and used the beaker containing the last of the mixture as a receptacle for coffee drugs. The reaction immediately occurred and he obtained a good yield of the required product". *Conclusion:* "The caffeine in the coffee was a catalyst for the reaction".

TABLE 2

Initial results

Form	Mean (/42)	S.D.	Pearson	Reliability (α)
C	26.1	4.89	0.28	0.64
P	27.1	4.67		0.67

N = 45

true, whichever test has been taken first. For such short tests, reliability coefficients are reasonable, but even if uncertainty due to small student numbers is borne in mind, the inter-test correlation can only be described as low.

Investigation showed that all but one or two students were studying both physics and chemistry. The items had been "vetted" and pronounced apparently satisfactory by a number of science educators, and most discriminated satisfactorily in practice. No comparison with correlation data between Forms Ym and Zm of the WGCTA could be made – this important aspect is not covered by the mass of data contained in the test manual.

Two further experiments were carried out in 1976. In the first, each of the three tests was split into two parts, one containing three subtests and the other two. To each part of Form C was added the corresponding parts of Form P and Form Ym, giving two tests of 85 and 99 items. In this way it was hoped to determine the degree of correlation between science and non-science items of the same type.

The figures are subject to uncertainty due to the small number of available students, but some points of interest are contained in the correlation results in Table 3.

TABLE 3

Critical Thinking Sub-Tests. Watson-Glaser Comparisons

C = Form C   P = Form P   W = Watson-Glaser

R = Inference sub-test
A = Assumptions "
D = Deduction "
I = Interpretation "
E = Evaluation "

	WR	PR
CR	0.45	0.34
WR		0.22

	WA	PA
CA	0.56	0.01
WA		0.42

	WD	PD
CD	0.39	0.16
WD		0.48

	WI	PI
CI	0.69	0.74
WI		0.45

N = 23

	WE	PE
CE	-0.15	-0.20
WE		0.40

Indications are that some subtests can more easily than others be made to give fairly consistent results. Thus interpretation (I) items are markedly consistent, but evaluation items much less so. Here the chemistry version is at odds with the others, and it can be seen that in 3 subtests out of 5, the worst correlation is between chemistry and physics items. Again practically all students had studied chemistry and physics.

The second 1976 experiment served to trial revised items for Form P and to provide more data for a comparison between Form C and P. Data were now available from 120 students.

As part of the major evaluation project mentioned earlier, data on Form C were available for a total of 243 cases. In addition, scores on the public examinations in chemistry and physics and on tests of chemistry knowledge, awareness of chemistry-society issues, problem-solving and the nature of science (a shortened version of TOUS) were available for correlation determinations. These are shown in a later table.

For Forms C and P, whole- and sub-test reliabilities and test and subtest intercorrelations are shown in Table 4.

TABLE 4

Combined data

	CSC	PSC	CR	CA	CD	CI	CE	PR	PA	PD	PI	PE
CSC (39)	(N=243) 0.67	.36	.54	.53	.63	.63	.58					
PSC (33)		.67						.72	.56	.76	.70	.75
CR (7)			.41									
CA (7)				.41								
CD (8)					.37							
CI (9)				Mean 0.41	=	.39						
CE (8)							.45					
PR (5)			.15					.31				
PA (5)				.13					.27			
PD (7)			Mean 0.11	=	.04			Mean	=	.34		
PI (5)						.14			0.39		.53	
PE (8)							.09					.36

N = 11

CSC, PSC = Total Scores ( ) = No. of items in subtest

Some items in each tests were judged unsatisfactory and are omitted from all scores.

The major diagonal shows reliabilities (α), the other figures being Pearson correlations. For such small numbers of items, Form C subtests show a non-negligible degree of internal consistency – the inter-subtest correlations are much lower. However, it is doubtful if, for example, a "Test of Evaluation of Arguments" would be useful by itself if based on this format. Watson and Glaser do not recommend using subtest scores because of low reliability coefficients (0.40-0.55 for Form Zm). Clearly, the five types of items devised by Watson and Glaser do not in fact measure distinct skills, nor has it been claimed that they should. Rust (1960) and Follmann (1969) have shown by factor analysis that the ability to do critical thinking tests is a composite of several skills.

The correlation between scores on Forms C and P is still low, and is comparable with the examination correlations shown later in Table 6. The inter-subtest correlations are particularly low. Clearly, knowledge of subject matter is not a major factor. Does the low correlation therefore imply that critical thinking in physics is different from critical thinking in chemistry?

As stated earlier, the tests were not initially intended to be strictly chemistry- or physics-based. In fact, there were two items on Form C which dealt with physics and one on Form P which dealt with chemistry. A number of items were science-based but not specifically chemistry or physics and some, chiefly deduction items, were only incidentally connected with science, and could have been re-worded to involve almost any subject.

Results for the two forms were combined (producing, incidentally, a "Physical Science Critical Thinking Test" of 70 items, $\alpha = 0.80$) and items were grouped as chemistry (C), physics (P), science (S) and general (G). These four groups were treated as new subtests. Reliabilities and correlations are shown in Table 5.

TABLE 5

Rearrangement of items

	C	P	P.E.B. Chemistry	P.E.B. Physics	S	G
C (23)	0.53	0.41	0.39	0.30	} 0.59	
P (20)		0.52	0.39	0.36		
S (12)	0.27	0.29	0.31	0.28	0.26	0.55
G (15)	0.34	0.50				0.52

The fact that items in a group were taken on different days renders this treatment suspect, but the results are at least worth comment.

For such short tests, internal consistency for C, P and G items is moderately good, but low for S items. C and P items correlate rather better than do the whole tests, but the coefficient seems to be rather lower than that for internal consistency. The data concerning correlations with other tests and examinations are given in Table 6.

It is likely that all scores are influenced by a common factor, presumably IQ. It would appear, therefore, that Form C measures a different ability from those involved in the other tests.

Conclusions

It would be premature to state firm conclusions based on the data so far obtained. The number of cases is not large, and data is being collected from a larger number in order to minimise uncertainty in the calculated correlations.

TABLE 6

Critical Thinking Tests — Correlations with other tests

	Chemistry Knowledge	TOUS	Chemistry-Impact on Society	Problem Solving	P.E.B. Chemistry	P.E.B. Physics
C	0.37	0.52	0.35	0.48	0.29	0.35
P					0.28	0.22

However, it would appear that workers attempting to devise tests of critical thinking based on the WGCTA approach must be aware of problems of validity arising from the choice of subject matter. Just what effect the subjects has on a person's ability to correctly assess statements cannot yet be described; the field is a practically untouched and probably difficult area of research.

References

ENNIS, R.H. A concept of critical thinking. *Harvard Educational Review*, 1962, 32, 82-83.
FOLLMAN, J. Factor analysis of three critical thinking tests, one logical reasoning test and one English test. *Yearbook of the National Reading Conference*, 1969, 8, 154-160.
RUST, V.I. Factor analysis of three tests of critical thinking. *Journal of Experimental Education*, 1960, 29, 177-182.
WATSON, G. & GLASER, E.M. *Watson-Glaser critical thinking appraisal.* New York: Harcourt, Brace & World, Inc., 1964.

Research in Science Education, 1977, 7, 139-148.

LABORATORY BASED TEACHING IN PHYSICAL SCIENCE AND ITS RELATION TO PUPIL ATTAINMENT

P. P. Lynch

Introduction

Within the last two decades there has been what might be termed a revolution in science education which has strongly influenced both teaching objectives and instructional methods. This revolution is essentially neo-heuristic with the emphasis on *guided* discovery, and in this respect is more controlled and more controllable than Armstrong's heurism (Herman, 1969). Support for the adoption of the discovery has been further stimulated by the compelling writings of behaviourists (Ausubel, 1968; Bruner, 1961; Gagné, 1965). Such a method is usually seen, ideally, as *replacing* conventional laboratory and traditional expository methods rather than an alternative. Yet, in spite of the virtually unreserved promotion of such methods in schools there is little if any evidence to justify their *teaching and learning* effectiveness over traditional methods (Babikian, 1971).

In more recent years discovery methods and related curricula have been introduced, to the Third World, where *cost-effectiveness* is of critical importance. It should be noted that most curriculum work in the Third World involves English as the medium of instruction. Inevitably, this curriculum work is neo-heuristic and, unfortunately, the much broader spectrum of European opinion is largely ignored. European opinion on such matters as teaching methods and the value of experimental work in schools varies widely — contrary to popular opinion (Thompson, 1972). We are *not* all moving in the same direction.

The problem

The results reported here have been abstracted from a national survey into the nature, purpose and organisation of practical work (Lynch, 1977), carried out in South African High Schools and Universities. Inevitably, within the general data available there is much information related to specific problems, in this case, the effectiveness of different teaching methods in relation to pupil attainment. The South African situation is of considerable interest because as far as science teaching is concerned teaching methods are influenced as much by European as British/American tradition. The Afrikaans speaking community draws much of its inspiration from German-Dutch tradition and this is maintained by the interchange of teachers and educators from those countries. The English speaking community tends, by comparison, to look to Britain and the U.S. for its teaching models. However, all school pupils complete the same matriculation examinations irrespective of language of instruction or teaching method employed.

Procedure and results

First year university students were asked to indicate their experience of *experimental techniques, specific types of practical work* and *standard experiments* at High School level by means of a questionnaire.

The questionnaire was constructed after a careful literature survey of similar studies (Kerr, 1963) and with the aid of an inquiry team and an advisory steering committee composed of 32 persons actively involved with school and university teaching in South Africa. The inquiry team and the committee reviewed each question and item with regard to syntax and relevance, and eliminated or added items. By this means the questionnaire was reviewed and rewritten three times, and we feel that full use was made of the collective wisdom and experience of all concerned.

The questionnaire was given to the students during a normal but previously undisclosed lecture period at the beginning of their specific first year courses. Students were given a short standardised preamble explaining the purpose of the survey, and aimed at winning the students' interest, cooperation and confidence. Each question was administered separately, clearly read and explained with an example. Approximately 40 minutes was taken to complete the items and by these means questionnaires were usually returned fully completed. A section for free comment was available and students were encouraged to use this. Careful scrutiny of this and the care taken with completion of individual items was taken as a measure of student cooperation. Less than 1% of the sample needed to be eliminated as a consequence of any tell-table indications of careless completion or facetious comment.

The items included in the questionnaire were as follows:

(a) *Frequency of use of experimental techniques:* A list of experimental techniques considered to be of central importance in secondary level physical science was drawn up with the help of the advisory steering committee. Of the fifteen techniques finally listed (Table 2), items 1, 2, 5, 9 and 10 are general techniques, items 3, 4, 6, 7 and 8 more specific to chemistry and items 11, 12, 13, 14 and 15 are more specific to physics.

Students were asked to indicate whether or not they *knew* about, had *seen done*, or had *carried out* the techniques themselves.

If the students marked the last category, they were asked to indicate the frequency using the 5 point scale

1 = never (but seen done)
2 = seldom
3 = moderately often
4 = often
5 = very often

The mean scores for each item are given in Table 2.

(b) *Frequency of use of different types of practical work.* Students were asked to indicate the extent to which *three* types of practical work were carried out at school, namely, *demonstrations, group work* or *individual work*. For the purpose of this inquiry and on the advice of the advisory steering committee it was decided that experiments involving 1 or 2 pupils would be termed 'individual work'. No attempt was made at this stage to distinguish between standard exercises, discovery experiments and project work.

The frequency of use of the three categories of practical work were measured on the 5 point scale

every period = 1
x 2 per week = 2
x 1 per week = 3
x 1 per month = 4
never = 5

Students were asked to indicate which category *most nearly* represented their High School experience.

The mean scores obtained are given in Table 3.

(c) *The use of standard experiments in physics and chemistry:* Fourteen experiments (seven in physics and seven in chemistry) taken from relevant school syllabi were considered by the advisory steering committee to represent a comprehensive coverage for any physical science syllabus (Table 4 below). The method of use or not of the standard experiments was measured in the categories 'know about (only)', 'seen done' and 'carried out myself'. The first category indicates no school experience of that particular standard experiment. The second category, 'seen done' indicates experience of a demonstration of the experiment. The third column, 'carried out myself' indicates direct experience in the form of individual or group work.

Mean scores based on the corresponding 3 point scale are given in Table 4. A discussion dealing with the percentages involved in the different categories has appeared elsewhere (Colussi, 1975).

The sample

The questionnaire responses for first year University students in Faculties of Science, Engineering and Medicine were obtained from an urban and a rural university in which the medium of instruction is specifically English (the Witwatersrand and Rhodes), and specifically Afrikaans (Pretoria and Stellenbosch). This total sample was considered to be representative of first year science, medical and engineering students at universities for white pupils in South Africa. Further details of the sample are provided in Table 1.

TABLE 1

Details of sample size and structure

University	Sample Size	Faculties Represented
Witwatersrand	728	Science/Engineering/Medicine
Rhodes	114	Science/Engineering
Pretoria	906	Science/Engineering/Medicine
Stellenbosch	604	Science/Engineering/Medicine
TOTAL	2352	

Analysis and display of results

The data from the questionnaire was transferred onto punch cards, verified, and finally stored on magnetic tape.

Analysis of the data was achieved using the computer package SPSS (Statistical Package for the Social Sciences). Mean scores, standard deviations and variances with regard to

TABLE 2

Mean scores for frequency of use of experimental techniques by students associated with different Examination Boards

Experimental Techniques	Entire Population	TVE	JMB	NSC	CAPE SC	OFSLC	NATAL SC	Other	Significance of differences
1. Weighing (on a rough balance	2.2	1.9	3.0†	2.2	2.1	2.4	3.3	3.1†	•
2. Weighing (to at least 3 figure accuracy)	1.8	1.4	2.5	1.5	1.8	2.7	2.6	3.2†	•
3. Titrating with a burette	2.0	1.6	2.9	1.4	2.1	2.4	3.4†	3.6†	•
4. Using a pipette	2.1	1.8	3.0†	1.8	2.1	2.7	3.4†	3.6†	•
5. Measuring of volume with a measuring cylinder	2.5	2.2	3.2†	2.2	2.4	3.1†	3.5†	3.4†	•
6. Filtering, using filter paper in a funnel	2.1	1.8	3.0†	1.8	2.1	2.6	2.9	3.1†	•
7. Crystallising	1.7	1.6	2.1	1.5	1.6	1.8	2.1	2.1	•
8. Distilling	1.4	1.3	1.7	1.3	1.4	1.5	1.8	1.9	•
9. Heating with a bunsen burner	2.9	2.4	4.1†	2.6	2.8	3.6†	4.3†	3.9†	•
10. Measuring with a metre rule	3.2†	3.0†	3.7†	3.1†	3.1†	3.9†	3.9†	3.7†	•
11. Measuring with a micrometer screw	1.4	1.3	2.0	2.2	1.3	1.4	1.4	2.3	•
12. Measuring pressure with a manometer or pressure gauge	1.4	1.3	1.6	1.5	1.3	2.0	2.0	1.8	•
13. Reading an ammeter or voltmeter	2.7	2.4	3.2†	3.9†	2.4	3.9†	2.8	3.0†	•
14. Reading a thermometer	3.0†	2.7	3.8†	2.8	2.9	3.8†	4.1†	3.9†	•
15. Connecting up an electrical circuit	2.4	2.1	3.0†	2.6	2.2	3.7†	3.6†	2.8	•
OVERALL MEAN SCORE	2.2	1.9	2.9	2.1	2.1	2.8	3.0	3.0	•

Note: † = mean score in moderate-often range.

a mean score less than 1.8 indicates that the majority have never carried out that technique

Key :

1	2	3	4	5
never	seldom	moderately often	often	very often

• = differences significant at the 0.5% level

TABLE 3

Mean scores for type of practicl work used by students associated with different Examination Boards

Type of Practical Work	Entire Population	TVE	JMB	NSC	CAPE SC	OFSLC	NATAL SC	Other	Significance of differences
Frequency of demonstrations	2.9	3.2	2.5	3.0	2.5	2.3	2.4	2.8	*
Frequency of group work	3.9	4.1	3.6	4.0	3.9	3.0	3.3	3.7	*
Frequency of individual work	4.4	4.6	3.7	4.4	4.3	3.9	3.7	3.4	*
Number of students in sample	2279	683	180	89	488	106	89	81	

Key :

1	2	3	4	5
every period	X2 per week	X1 per week	X1 per month	never

* = difference significant at the 0.5% level

TABLE 4

Mean scores for use of standard experiments by students associated with different Examination Boards

Standard Experiments	Entire Population	TVE	JMB	NSC	CAPE SC	OFSLC	NATAL SC	Other	Significance of differences
PHYSICS									
1. 'Acceleration'	2.2	2.1	1.9	1.9	2.5	2.6	2.3	$1.5^{\#}$	•
2. 'g'	2.1	1.8	1.9	$1.7^{\#}$	2.6	2.5	2.1	$1.7^{\#}$	•
3. 'Specific or latent heat'	$1.7^{\#}$	$1.5^{\#}$	1.8	1.8	1.9	2.2	2.1	$1.5^{\#}$	•
4. 'Focal length'	1.9	1.9	1.8	1.9	2.0	1.9	2.0	$1.6^{\#}$	
5. 'Electrical resistance'	1.9	1.9	1.9	2.0	1.9	2.7	2.3	$1.6^{\#}$	•
6. 'Electric motor'	1.8	$1.7^{\#}$	1.8	2.1	1.8	2.0	1.8	$1.4^{\#}$	•
7. 'Interference patterns'	2.0	2.0	1.9	2.1	2.2	2.5	1.9	$1.7^{\#}$	•
CHEMISTRY									
1. 'Solubility'	1.9	$1.7^{\#}$	2.0	1.9	2.0	2.2	2.0	$1.6^{\#}$	•
2. 'Chemistry of an element'	$1.6^{\#}$	$1.5^{\#}$	1.8	$1.6^{\#}$	$1.7^{\#}$	$1.7^{\#}$	1.9	$1.5^{\#}$	•
3. 'Organic or inorganic prep'	$1.7^{\#}$	$1.6^{\#}$	1.8	$1.5^{\#}$	1.8	1.9	1.8	$1.6^{\#}$	•
4. 'Rate of reaction'	$1.7^{\#}$	$1.6^{\#}$	$1.7^{\#}$	$1.7^{\#}$	1.8	2.0	2.2	$1.4^{\#}$	•
5. 'Molecular models'	1.9	1.8	1.8	1.9	2.1	1.8	1.8	$1.6^{\#}$	•
6. 'Vapour density or molecular weight'	$1.3^{\#}$	$1.3^{\#}$	$1.4^{\#}$	$1.3^{\#}$	$1.3^{\#}$	$1.7^{\#}$	$1.6^{\#}$	$1.5^{\#}$	•
7. 'Acids and bases'	1.8	$1.7^{\#}$	2.0	1.8	2.1	2.2	2.2	$1.6^{\#}$	•

= majority of students have neither seen the experiment nor carried it out themselves.
a mean score greater than 2.4 indicates that the majority have carried out that experiment themselves.

Key:	1	2	3
	know about (only)	seen done	carried out myself

• = differences significant at the 0.5% level individual items were calculated for total samples and for various sub-groups. The significances of the differences were assessed using one way analyses of variances (F tests). *In the tabulated data, here, the level of significance of the difference if it exceeds the 0.5% level is marked thus,* *. This relatively high level of significance has been used in view of the inherent difficulties that are associated with the interpretation of data obtained by survey methods (Gardner, 1975).

Findings – General

Students indicate that the choice of content, the teaching method and the time spent on practical works at High School varies widely. Student experience of even very elementary practical techniques such as heating with a bunsen burner or connecting up an electric circuit varies considerably. A full analysis and description of this experience and its implications are provided elsewhere (Colussi, 1976).

In order to explore the factors most important in determining these differences the sample was examined with respect to variables such as sex differences, type and size of schools, language medium and Examination Board.

There are some considerable differences in experience (judged by mean scores) between private and government schools, between single sex and co-educational schools and between English and Afrikaans medium schools. However the greatest overall differences between mean scores are produced when the sample is examined with respect to Examination Board.[1]

Mean scores for the entire population and for the different examination boards are given for that reason. The scale used in Table 2, is non-linear but as a rough guide OFS and Natal SC students indicate experience, on average, of twice as many demonstrations and almost four times as much group work as TVE students. By comparison with the 'entire population', OFS, Natal SC, JMB and 'Other' students have considerably more experience of all three types of practical work.

The results in Table 3 are consistent with those obtained in Table 2. The overall mean score gives a crude indication of the average frequency of techniques and OFS, Natal SC, JMB and 'Other' students have a mean score equivalent to 'moderately often'. Students associated with the TVE, NSC and Cape SC have mean scores approximately equivalent to 'seldom'.

The scale used in Table 4 is particularly constricted. A mean score less than 1.8 is associated with the majority of students having no experience of that experiment, while a mean score greater is associated with the majority having experience of individual or group work.

The results for the 'Other' group suggest that the latter have a more technique oriented background than South African students. The TVE column in Table 4 needs to be interpreted with caution since further examination showed that English and Afrikaans medium schools have very different experience of the standard experiments listed.

The prime reason for presenting the data in this form is to show that the experience of the items chosen in the questionnaire do, indeed, vary widely, and that there are significant differences between certain sub-groups.

Findings with respect to matriculation grade

The mean scores for students with different matriculation grades are given in Tables 5-7. The codification, A to F represents grades from distinction (A) to fail (F).

TABLE 5

Mean scores for frequency of use of experimental techniques by students with different matriculation grades

Experimental Techniques		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Matriculation Grade	A	2.1	1.6	1.9	2.1	2.3	2.0	1.6	1.3	2.7	3.2	1.3	1.4	2.6	2.9	2.4
	B	2.1	1.8	1.9	2.0	2.6	2.0	1.6	1.3	2.8	3.3	1.3	1.5	2.7	3.0	2.4
	C	2.1	1.7	2.0	2.1	2.5	2.1	1.7	1.4	2.8	3.1	1.4	1.4	2.7	2.9	2.4
	D	2.1	1.7	1.9	2.1	2.5	2.1	1.7	1.4	2.8	3.1	1.4	1.3	2.6	2.9	2.3
	E	2.2	1.9	1.8	2.0	2.4	2.0	1.6	1.5	2.6	3.1	1.6	1.3	2.5	2.9	2.2
	F	2.1	1.9	2.1	2.3	2.6	2.2	2.0	1.5	3.5	3.2	1.3	1.1	2.7	3.0	2.4
Significance of differences			•					•		•						

TABLE 6

Mean scores for type of practical work used by students with different matriculation grades

Type of Practical work		Frequency of demonstrations	Frequency of group work	Frequency of individual work
Matriculation Grade	A	3.0	4.0	4.5
	B	2.9	3.9	4.4
	C	2.8	3.8	4.4
	D	2.9	3.9	4.4
	E	2.9	3.9	4.4
Significance of differences				

Students with F (fail) grades at Matriculation level have rather less experience of practical work but the differences are not usually significant at the 0.5% level.

There are a few instances where the differences of experience are significant (see Tables 6 and 7) and these are due to F grade students. But the similarities are more striking than the differences particularly when compared with the previous tables.

Our interpretation of these results is that there is an absence of any correlation between amount of practical work, type or nature of practical work experienced and matriculation grade obtained.

TABLE 7

Mean scores for frequency of use of standard experiments by students with different matriculation grades

Physics experiments	Physics Experiments							Chemistry Experiments						
	1	2	3	4	5	6	7	1	2	3	4	5	6	7
A	2.3	2.2	1.7	1.8	1.8	1.8	2.1	1.9	1.7	1.7	1.7	1.9	1.3	1.9
B	2.2	2.1	1.7	1.9	1.9	1.8	2.1	1.8	1.6	1.7	1.7	1.9	1.3	1.8
C	2.1	2.0	1.7	1.9	1.9	1.8	2.0	1.8	1.6	1.6	1.7	1.8	1.4	1.8
D	2.2	2.1	1.7	2.1	2.0	1.8	2.1	1.9	1.6	1.7	1.7	1.8	1.3	1.8
E	2.3	2.0	1.7	2.0	2.0	1.8	2.1	1.9	1.6	1.7	1.8	1.8	1.4	1.8
F	2.3	2.1	2.0	2.0	2.0	2.0	2.2	2.0	1.4	1.6	1.7	2.0	1.3	1.7
Significance of differences			*						*					

Discussion

The terminal examination in question, University Matriculation, is little different to the equivalent examinations used in, for example, the U.K., the U.S. or Australia. Use is made of both multiple choice, structured, and essay type questions and every attempt is made to cover the usual accepted range of cognitive skills. The syllabus content, itself, shows the influence of various well established curricula such as Chem Study, PSSC Physics, Nuffield O-level and A-level etc. The standard of attainment aimed at is about equivalent to one year of sixth form work in the U.K. It is performance, measured by this examination, which is unrelated to choice or even use of practical work. There is, thus, no evidence that practical work *reinforces learning* or that it encourages students to *learn more* in relation to established 'academic' assessment.

Neither the motivational affects nor the acquisition of laboratory or related organisational skills are considered here but will be reported at a later stage. But, the absence of any apparent benefit as far as the acquisition of scientific knowledge is concerned has some serious implications for curriculum developers, generally. If cost-effectiveness is of a paramount importance as in third world situations, then even more attention needs to be paid to the nature and extent of the practical work used.

Discovery-type programmes are more expensive to set up and to maintain, apart from, in many cases, being more difficult to organise at a teaching, in-service and administrational level. Programmes placing more emphasis on the *lecture-demonstration method* may offer a more convenient, cheaper and as effective an answer.

Footnote

1. The codifications used in Tables 2, 3 and 4 refer to the Examination Boards, and their respective Matriculation Examinations required for university entrance.

They are as follows: TVE = Transvaal University Matriculation, JMB = Joint Matriculation Board, NSC = National Senior Certificate, Cape SC = Cape Senior Certificate, OFSLC = Orange Free State Leaving Certificate, Natal SC = Natal Senior Certificate.

'Other' = (usually) U.K. or Rhodesian O/A level Certificate.

References

AUSUBEL, D.P. *Educational psychology: a cognitive view.* New York: Holt, Rinehart and Winston, 1968.

BABIKIAN, Y. An empirical investigation to determine the relative effectiveness of discovery, laboratory and expository methods of teaching science concepts. *Journal of Research in Science Teaching*, 1971, 8(3), 201-209.

BRUNER, J.S. Act of discovery. *Harvard Educational Review*, 1961, 31, 21-32.

COLUSSI, M. Practical work in school science teaching. *Journal of Education, University of Natal* (South Africa), 1975, 7, 55-68.

COLUSSI, M. Practical work in science teaching in South African schools. Unpublished M.Ed. Thesis, University of the Witwatersrand, Johannesburg, South Africa, 1977.

GAGNÉ, R. *The conditions of learning.* New York: Holt, Rinehart and Winston, 1965.

GARDNER, P.L. Scales and statistics. *Review of Educational Research*, 1975, 45(1), 43-57.

HERMAN, G.D. Learning by discovery — a critical review of studies. *Journal of Experimental Education*, 1969, 38, 58-72.

KERR, J.F. *Practical work in school science.* U.K.: Leicester University Press, 1963.

LYNCH, P.P. The nature, purpose and organisation of experimental work in high schools and universities in South Africa. Human Sciences Research Council of South Africa (Pretoria), 1977.

THOMPSON, J.J. *European curriculum studies.* Strasbourg: Chemistry Council of Europe, 1972, No. 4.

Research in Science Education, 1977, 7, 149-156.

ATTITUDES TO SCIENCE : A SEMANTIC DIFFERENTIAL INSTRUMENT

Renato Schibeci

Introduction

Much research has been published in the area of attitudes to science. A number of comprehensive reviews of the literature have appeared: Aiken and Aiken (1969); Gardner (1975).

This paper is in two parts. The first part explores the distinction between *attitudes to science* and *scientific attitudes* which has been proposed by some workers. The ethical problems which can arise in the measurement of these two types of attitudes are examined. Some of these issues have not been sufficiently aired and debated. In particular, the implication for the classroom teacher have not been sufficiently investigated.

The second part of the paper describes the development of a semantic differential instrument to measure some aspects of attitudes to science. This instrument was developed in conjunction with the use of some innovative chemistry curriculum units.

The affective domain in science education

(a) A change in emphasis

Science curricula in the secondary schools of the different Australian states now place a greater emphasis on objectives in the affective domain than they did in the past. The change in the N.S.W. science curriculum for years 7-10 illustrates this change. The detailed syllabus (Secondary Schools Board, 1968) with its almost exclusive emphasis on cognitive objectives has been replaced by a much more flexible curriculum (Secondary Schools Board, 1973a) with a greater emphasis on affective objectives. The old syllabus had been assessed by an external examination dominated by cognitively-oriented items. The Secondary Schools Board (1973b) suggested that student progress in the three areas of the new curriculum be weighted as follows: interests and attitudes, up to thirty percent; knowledge and understanding, up to fifty percent; skills, up to twenty percent.

Bloom's (1956) taxonomy of educational objectives in the cognitive domain had an immediate impact on curriculum development. The taxonomy listed a number of cognitive behaviours which are amenable to evaluation. A variety of teaching strategies is available for producing these cognitive skills.

Krathwohl's (1964) taxonomy of affective objectives has not had the same impact as its predecessor. The affective domain in education does not offer the same safe, smooth sailing of the cognitive domain; reefs and shoals abound. It is important to explore the issues involved in evaluation in this turbulent domain.

(b) Two types of attitudes

Bloom (1971) cites three reasons for the past neglect of the affective domain in the U.S. Firstly, there is the fear that teachers will indoctrinate students; secondly, the time taken to induce attitude change is regarded as a serious problem; thirdly, there is the fear that teachers would invade the privacy of students. These fears reveal a confusion which exists about affective objectives.

Many meanings are given to "attitude" in science education. Aiken and Aiken (1969), in reviewing the literature on attitudes to science in the U.S., categorised studies into three areas: like vs. dislike of science (or a particular science); like vs. dislike of scientists and their activities; adherence to "scientific method".

Gardner (1975) proposed two broad categories. The first category is concerned with studies of "attitudes to science" involving a favourable or unfavourable attitude to an attitude object such as "science" or "scientists". The second category is concerned with studies of "scientific attitudes" involving traits (such as honesty, open-mindedness, and curiosity) which scientists are supposed to display.

Moore and Sutman (1970) proposed a similar distinction in which they divided "scientific attitudes" into two groups: intellectual and emotional. An example of an *intellectual* attitude is: "Observation of natural phenomena is the basis of scientific explanation". Intellectual attitudes are cognitively-oriented: they are based on knowledge of the psychological object of the attitude. An example of an *emotional* attitude is: "Being a scientist would be dull and uninteresting". Emotional attitudes are affectively-oriented: they are based on feelings or emotions towards the psychological object of the attitude.

The three concerns about affective outcomes outlined above (indoctrination, time taken to change attitudes and invasion of privacy), address themselves mainly to affectively-oriented (emotional), rather than to cognitively-oriented (intellectual) outcomes of science teaching. There can be no quarrel with the promotion of "scientific attitudes" if these are regarded by scientists as crucial to the scientific enterprise. Gauld (1976) has, however, raised some doubts about the place of "scientific attitudes" in science.

The promotion of certain "attitudes to science" is more contentious. It is at this point that the "invasion of privacy" issue becomes relevant. Does the teacher have any right to change students' commitments?

Many would be unhappy for a teacher to have access to a student's private emotions and feelings. There would be less cause for concern if teachers gathered group data on feelings towards science (as opposed to the feelings of an individual). The gathering of data on affectively-oriented "attitudes to science" would have to, in other words, ensure that the privacy of the individual is protected. If curriculum planners consider affectively-oriented attitudes as important outcomes, they must not expect data on individual students to be collected. A student must not be penalised because he does not display the "correct" attitude to science.

(c) Assessment of attitudes

Once the ethical problems associated with attempts to measure both types of attitudes are resolved, the problem of valid assessment methods remains.

The assessment of cognitively-oriented *scientific attitudes* has been attempted recently by Kozlow and Nay (1976). They developed a multiple-choice instrument to measure eight scientific attitudes: critical mindedness, suspended judgment, respect for evidence, honesty, objectivity, willingness to change opinions, open-mindedness and questioning attitude. A behavioural description of each of these attitudes is provided.

This kind of approach is appealing because of its simplicity and hence ready applicability in the classroom. Students may, of course, give the responses they thought the teacher desired. The validity of the test needs to be carefully established. This approach to the assessment of scientific attitudes remains the most fruitful to date.

A variety of strategies to measure *attitudes* has been developed in social psychology. The most popular technique is the use of Likert-type scales; most of the scales reported in Shaw and Wright (1967) are of this type. Other techniques include: forced-choice selection methods; paired comparisons method; method of equal-appearing intervals; method of summated rations; free response methods; and, semantic differential technique.

The methods described above have been used chiefly in the assessment of affectively-oriented attitudes to science. There are a number of theoretical and methodological issues associated with these techniques which pose difficulties in the classroom. The teacher must be willing to invest considerable time and effort in developing a valid measure to provide group data on the attitudes to science of the classes he teaches. There are no readily available "standardised" tests in this area.

The semantic differential instrument

(a) The Semantic Differential (SD) technique

The aim of this study was to develop a semantic differential instrument to measure some aspects of affectively-oriented attitudes (in the sense described in the previous section). This instrument was developed in order to help in the evaluation of the impact of some innovative chemistry curriculum units.

The SD is now firmly established as an attitude measurement technique as well as an important tool in psycholinguistic theory:

> "In social science, where measurement techniques are added and discarded from the armamentorium with amazing alacrity and where fashion often rules the paths of inquiry, it is gratifying that the semantic differential (SD) technique has so far outlived its initial rush of acceptance as a fad."
>
> (Osgood *et al*, 1975, p.37)

The semantic differential does not require the item analysis and revision procedures required by other techniques. If two instruments validly and reliably measure the same aspects of attitudes, the SD is to be preferred. For example, McCallon and Brown (1971) found that the Likert-type instrument ("Mathematics Attitude Scale") and an SD instrument they developed gave a correlation coefficient of $r = 0.90$.

(b) The Test Instrument

The SD instrument used in this study consisted of eight concepts (CHEMISTRY, LABORATORY WORK, SCIENCE, SCIENTISTS, SCIENCE TEACHER, SCHOOL, SCIENCE CLASS, SCIENCE LABORATORY). These concepts were agreed on after discussion with a group of science teachers.

The scales had been used in a study of attitudes to physics in the U.S. (Geis, cited in Klopfer, 1971). These scales were:

> "safe/dangerous, uninvolved/involved, enjoyable/unenjoyable, threatening/comforting, refreshing/weary, solemn/cheerful, organised/messy, unproductive/productive, complex/simple, important/unimportant, interesting/dull, gloomy/joyful, worthless/worthwhile, risky/sure, boring/exciting.

(c) Subjects

The population in the study consisted of year 10 science students in a Perth metropolitan high school. Students were asked *not* to write their names on the forms and were told that their teachers would not see the raw data.

The SD was administered in July, 1976 (n = 155) and re-administered in November, 1976 (n = 135). There was some reorganisation of classes between the two test occasions. Because the SD was filled in anonymously, it was difficult to ensure that exactly the same populations was involved on both occasions. Some students could not remember if they had taken the test on the first occasion!

(d) Factor analysis

The semantic differential produces a cube of data whose dimensions are P (people) X C (concepts) X S (scales); a three mode factor analysis (Tucker, 1966) would be the most attractive method of analysis. This procedure, which has been used by Reid (1969), is not widely available.

A computer program was written to transform the raw SD data. For example, a score of 3 on the uninvolved/involved scale became 5. The program thus produced scores as if all the positive adjectives of the bipolar adjective pair had been on the left hand side of each table. Factor analysis was carried out on a matrix produced by a "stringing out" procedure. P people rating C concepts on S scales results in a matrix with (P X C) rows by S columns (Maguire, 1973).

In this study, 155 people rated 8 concepts on 15 scales, giving a matrix with (155 X 8) rows and 15 columns.

The matrix was factor analysed using a program in the Statistical Package for the Social Sciences (Nie et al, 1975). The method selected was principal factoring with iteration (perhaps the most widely used factoring method) with the Varimax rotation.

(e) Results

The means and standard deviations for the fifteen scales are presented in Table 1. Table 2 shows the rotated factor loadings for each of the fifteen scales on both test occasions.

The three factors which emerged from the factor analysis have been labelled INTEREST, IMPORTANCE and SAFETY. The study by Geis (cited in Klopfer, 1971) had revealed four factors: INTEREST, IMPORTANCE, DIFFICULTY and SAFETY.

These scales (with the exception of simple/complex) appear to be suitable in measuring two aspects of attitudes to science (interest and importance). The scale simple/complex will not be used in future studies; results obtained on this scale do not contribute to measurement of attitudes.

(f) Evaluation of some chemistry curriculum units: a note

The SD instrument is being developed to provide some data on some chemistry curriculum units. The chemistry curriculum units were taken from the "Action Chemistry" programme developed in the U.S. by Bolton and others (1973). This programme is described as a student-centred laboratory programme in general chemistry. The materials are attractively

TABLE 1

Means and standard deviations for the fifteen scales

Scale	Mean	Standard Deviation
Safe-Dangerous	3.65 (3.63)	1.83 (1.73)
Involved-Uninvolved	2.52 (2.75)	1.64 (1.66)
Enjoyable-Unenjoyable	3.49 (3.48)	1.87 (1.69)
Comforting-Threatening	4.24 (4.14)	1.45 (1.36)
Refreshing-Weary	4.18 (4.09)	1.56 (1.48)
Cheerful-Solemn	4.14 (4.14)	1.63 (1.49)
Organised-Messy	3.05 (3.14)	1.92 (1.91)
Productive-Unproductive	2.77 (2.94)	1.67 (1.73)
Simple-Complex	4.90 (4.67)	1.77 (1.73)
Important-Unimportant	2.58 (2.73)	1.74 (1.72)
Interesting-Dull	3.15 (3.28)	1.83 (1.74)
Joyful-Gloomy	3.84 (3.92)	1.50 (1.39)
Worthwhile-Worthless	2.67 (2.86)	1.70 (1.69)
Sure-Risky	3.68 (3.75)	1.58 (1.47)
Exciting-Boring	3.75 (3.85)	1.74 (1.61)

Occasion 1 : n = 155
" 2 : n = 135 (in brackets)

TABLE 2

Rotated factor loadings

Scale	Factor 1	Factor II	Factor III
Exciting-Boring	0.73 (0.77)	0.29 (0.27)	0.16 (0.20)
Interesting-Dull	0.71 (0.74)	0.42 (0.35)	0.14 (0.20)
Joyful-Gloomy	0.71 (0.72)	0.16 (0.12)	0.25 (0.31)
Refreshing-Weary	0.68 (0.67)	0.14 (0.15)	0.20 (0.33)
Enjoyable-Unenjoyable	0.64 (0.69)	0.27 (0.24)	0.23 (0.24)
Cheerful-Solemn	0.63 (0.63)	-0.09 (-0.05)	0.27 (0.42)
Important-Unimportant	0.26 (0.45)	0.68 (0.66)	0.20 (0.12)
Worthwhile-Worthless	0.36 (0.45)	0.64 (0.63)	0.19 (0.15)
Productive-Unproductive	0.26 (0.35)	0.60 (0.61)	0.17 (0.24)
Involved-Uninvolved	0.10 (0.21)	0.50 (0.57)	0.02 (0.26)
Organised-Messy	0.16 (0.23)	0.49 (0.57)	0.22 (0.29)
Simple-Complex	0.23 (0.18)	-0.44 (-0.44)	0.05 (0.00)
Safe-Dangerous	0.19 (0.19)	0.11 (0.21)	0.66 (0.61)
Sure-Risky	0.19 (0.26)	0.21 (0.27)	0.57 (0.57)
Comforting-Threatening	0.42 (0.39)	0.07 (0.13)	0.53 (0.60)

Occasion 1 : n = 155
" 2 : n = 135 (in brackets)

presented, and designed to stimulate student interest in chemistry.

One of the classes in the sample used two of the chemistry units; another class used one unit. The semantic differential was given on two occasions. The factor loadings could be used to calculate INTEREST and IMPORTANCE scales. These showed that some changes in interest did occur, but it was not possible to determine whether this change was significant because of the reorganisation of the classes between the two test occasions.

There was some evidence, however, that the teacher was a much more significant variable than the curriculum variable. This is supported by other studies.

One of the major problems is to ensure that the new materials are used in the way the developers intended. Observation of the two classes using the materials indicated that this was not the case. Teachers at the school were, however, sufficiently impressed with the material to include some of these units in future Year 10 programmes.

Postscript

This paper has raised a number of issues related to the affective domain in science education. These issues need to be aired and debated. Curriculum developers who include attitude objectives in science programmes need to alert teachers to the normative and empirical issues involved.

The semantic differential technique provides a measure of some aspects of attitudes to science. Instruments of this type are useful in gathering data for curriculum evaluation.

The particular requirement in N.S.W. for science teachers to add together marks in the knowledge, attitude and skills areas needs to be seriously reviewed.

References

AIKEN, L.R. & AIKEN, D.R. Recent research on attitudes concerning science. *Science Education*, 1969, **53**, 295-230.

BLOOM, B.S. (Ed.) *Taxonomy of educational objectives: The classification of educational goals. Handbook 1. Cognitive domain.* N.Y.: McKay, 1956.

BLOOM, B.S., HASTINGS, J.T. & MADAUS, G.F. *Handbook on formative and summative evaluation of student learning*. N.Y.: McGraw-Hill, 1971.

BOLTON, RUTH P., LAMPHERE, E.V., MENESINI, M. & HUANG, P.C. *Action chemistry: Matter-energy-change: Teacher's guide.* N.Y.: Holt, Rinehart & Winston, 1973.

GARDNER, P.L. Attitudes to science: A review. *Studies in Science Education*, 1975, **2**, 1-41.

GAULD, C.F. Scientific attitudes and personal development. *Australian Science Teachers' Journal*, 1976, **22**, 73-82.

KLOPFER, L.E. Evaluation of learning in science. In Bloom, B.S. et al (eds.) *Handbook on formative and summative evaluation of student learning*. N.Y.: McGraw-Hill, 1971, 559-642.

KOZLOW, M. & NAY, A. An approach to measuring scientific attitudes. *Science Education*, 1975, **60**, 137-172.

KRATHWOHL, D.R. (Ed.) *Taxonomy of educational objectives: The classification of educational goals. Handbook II. Affective domain.* N.Y.: Mackay, 1964.

MAGUIRE, T.O. Semantic differential methodology for the structuring of attitudes. *American Educational Research Journal*, 1973, **10**, 295-306.

McCALLON, E.L. & BROWN, J.D. A semantic differential instrument for measuring attitude towards mathematics. *Journal of Experimental Education*, 1971, **39**, 69-72.

MOORE, R.W. & SUTMAN, F.X. The development, field test and validation of an inventory of scientific attitudes. *Journal of Research in Science Teaching*, 1970, **7**, 85-94.

NIE, N.H. *Statistical package for the social sciences.* (2nd ed.). N.Y.: McGraw-Hill, 1975.

OSGOOD, C.E., MAY, W.H. & MIRON, M.S. *Cross-cultural universals of affective meanings.* Urbana: University of Illinois Press, 1975.

REID, J.S. A three-mode factor analysis of students' perceptions of a university. *Journal of Experimental Education*, 1969, **38**, 93-96.

SECONDARY SCHOOLS BOARD. *Syllabus in science ordinary and advanced levels, Forms I-IV.* Sydney: Government Printer, 1968.

SECONDARY SCHOOLS BOARD. *Revised science syllabus Forms I-IV. Statement of aims and content.* Sydney: Government Printer, 1973a.

SECONDARY SCHOOLS BOARD. *Syllabus in science Forms I-IV Assessment guide.* Sydney: Government Printer, 1973b.

SHAW, M.E. & WRIGHT, J.M. *Scales for the measurement of attitudes.* N.Y.: McGraw-Hill, 1967.

TUCKER, L.R. Some mathematical notes on three-modes factor analysis. *Psychometrika*, 1966, **31**, 279-311.

Research in Science Education, 1977, 7, 157-164.

SUBJECTIVELY ASSESSED STUDENT PROFILES IN PHYSICS

Roger J. Osborne

Introduction

Teachers acquire a considerable store of knowledge about their students by observing a wide variety of student responses, as Harlen (1975) and others have noted. The behaviour of a student in the formal teaching situation, in class discussion and in the laboratory, and his or her written responses to a variety of assignments and tests, all affect the teachers' view of the student. The possible usefulness of this subjective information, for example for guidance or selection, is worthy of consideration. This would appear particularly so with the decrease in emphasis on examinations and the trends towards the assessment of students with respect to a wider variety of objectives. This paper considers one way in which this subjective information can be categorized and quantified, and explores its potential usefulness in one particular educational system.

General learner objectives have been developed to orientate and diversify teaching, learning and assessment for some physics courses in New Zealand. At the secondary school seventh form level (Bursary Physics) a set of Klopfer based objectives (Klopfer, 1971) is enunciated in the nationally available Teachers Guide (see C.D.U., 1973). At the University of Waikato, 21 learner objectives have been explicitely stated for the first year physics course (see Osborne and Freyberg, 1977). Without undue distortion of the objectives of either course it is possible to identify six major groupings of objectives common to both courses. The headings for these groups are listed in Table 1 (P.1 - P.6) along with a brief discussion about each heading, which was used in this study to guide teachers, in the selection of suitable criteria on which to base subjective assessments of student competence.

The study

The student subjects who formed the basis of this study were those first year University physics students at the University of Waikato in 1973 and 1974 who had completed the Bursary Physics course (seventh form) in New Zealand schools in the preceding year (N = 101 in 1973, N = 87 in 1974). Their former physics teachers (N = 50) were asked in 1973 and 1974 to subjectively assess their students on the six major aspects of physics instruction previously mentioned, and also on three more general aspects of academic ability and motivation (see Table 1, U.1 - U.3). A seven point scale was used to assess each criterion*, e.g. Comprehension of Physics HIGH: — : — : — : — : — : — : — : LOW. Teacher co-operation was such that nine subjective assessments were obtained for all 188 students.

In conjunction with this study, teachers were also asked what importance they considered should be given to each of the "physics objectives" in *teaching* at, (A) the seventh form level and, (B) the University level. For each objective they were asked to choose between Very Important, Important, Possibly Important, Not Important, Undesirable. Two teachers refrained from commenting on University teaching, no teacher considered any objective undesirable, and the percentage of teachers who chose each Importance category is given in Fig. 1.

---

*Teachers were asked to compare, as far as they were able, their students with the national set assuming the following national percentage distribution: HIGH : 5 : 10 : 20 : 30 : 20 : 10 : 5 : LOW.

TABLE 1

Subjective assessment criteria guide for teachers

Knowledge and Ability in Physics

P1 *Comprehension of the Concepts in 7 form Physics* – an understanding of the basic ideas of physics as presented by direct teaching, or from textbooks. A student low on this scale would have little useful knowledge of physics and probably a considerable number of misconceptions, e.g. understanding of centripetal force and acceleration, Gauss Law, Ampere's Law, etc.

P2 *Ability to apply physics to simple standard problems* – a student's ability to apply theory to problem situations, e.g. an ability to do homework problems and to solve application problems in tests and exams. (Although this ability is not independent of comprehension of physical ideas, some students show a particular facility for this aspect of physics despite some misconceptions about physics while alternatively some students comprehend the ideas but have difficulty applying them.)

P3 *Ability to critically analyse and evaluate ideas in physics* – a student's ability to think critically, to realize the full implications of derivations, and to critically analyse apparently satsifactory statements in physics, e.g. a student's critical study of material presented to him, and the perceptive enquiry of points which are not clearly presented.

P4 *Ability to carry out experimental work* – a student's ability in the laboratory, e.g. ability to set up an experiment, interpret instructions in terms of equipment provided, practical ability, etc.

P5 *Ability to process experimental data and draw conclusions* – includes the analysis of graphs, interpretation of results, and drawing conclusions from the results. A measure of this scale would be the work in a student's laboratory notebook particularly as it relates to detailed analysis of results and conclusions.

P6 *Positive attitude to physics* – a student's apparent satisfaction with school physics, e.g. a teacher's subjective impression as to whether the pupil found physics enjoyable and satisfying.

Suitability for Academic work in General

U1 *Academic ability* – a student's general intellectual ability with regard to 7 form work in so far as it can be separated from his scholastic achievement, e.g. reasoning ability, ability to ask intelligent questions and provide thoughtful answers to questions.

U2 *Academic Industry* – a student's academic industry in the 7 form, e.g. effort put into homework, study habits, attitude toward study, questioning of doubtful points, etc.

U3 *Maturity/Stability* – this scale indicates a student's independence and stability of purpose, A student high on this scale would not be unduly influenced by other students or affected by change to a new environment, e.g. university. The scale is independent of academic ability or the level of purpose (e.g. low or high achievement aims).

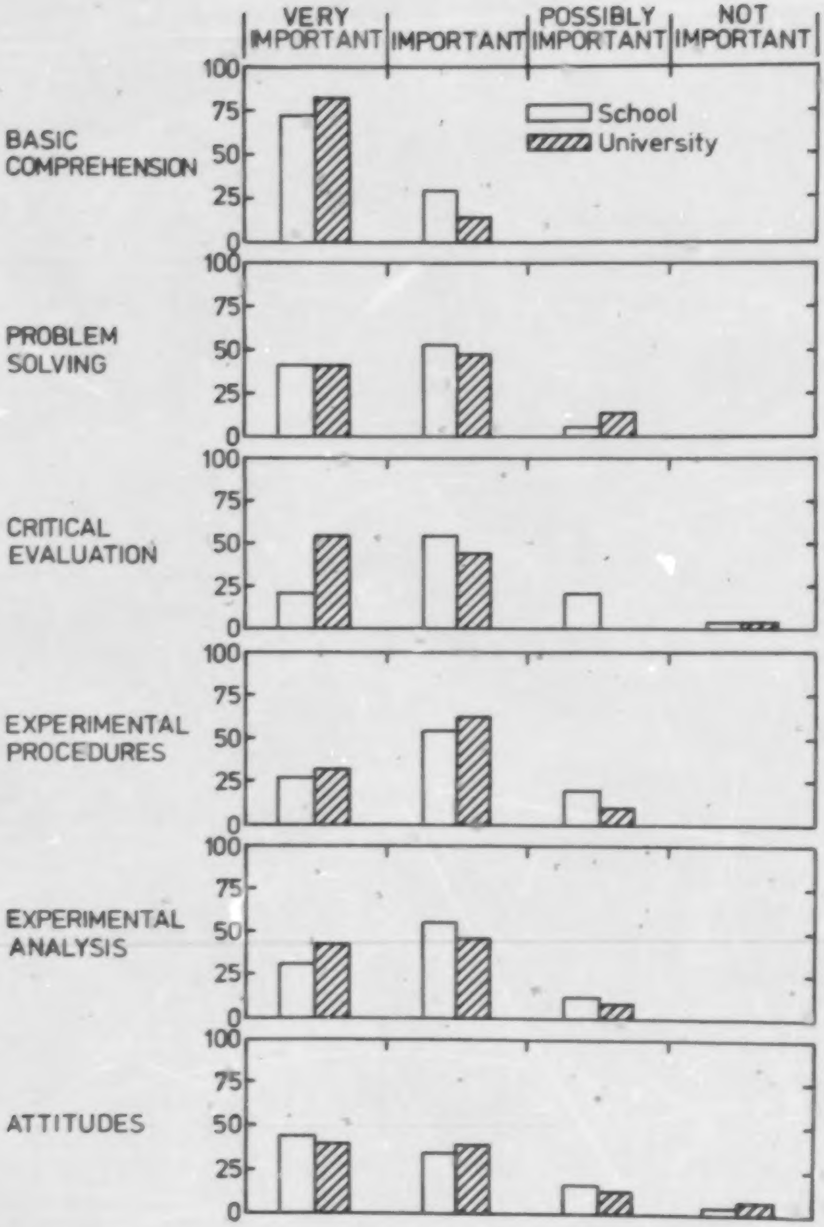


FIGURE 1

Teachers' views of the relative importance in the teaching of physics at school and university, of various physics education objectives.

In general, all objectives were considered important. Basic comprehension was considered of prime importance both at school and university level but critical evaluation of ideas was considered distinctly more important for university level teaching than for the school level. There was considerable divergence of opinion over the importance of attitudes in teaching physics. Some teachers took the view "teach them physics and let the attitudes look after themselves".

The analysis of subjective student assessments

To obtain an indication of the teachers' view of student attainment a principal components analysis of teachers' assessments on the nine scales of Table 1 was carried out for each year separately. In both cases, two significant factors were obtained (unrotated eigen values $\geqslant 1.0$). A Varimax rotation of these two factors resulted in the factor loadings shown in Table 2. Factor I is clearly associated with a student's "academic ability" as perceived by the physics teachers and Factor II could be interpreted as closely related to the teachers' perception of the students' "motivation towards academic work".

TABLE 2

Principal components analyses of teacher's subjective assessments of first year physics students in 1973 (N = 101) and in 1974 (N = 87) with factor loading >0.5.

Objective/Criteria	Factor				Mean 0-6 Scale		Standard Deviation		$100h^2$	
	I		II							
	1973	1974	1973	1974	1973	1974	1973	1974	1973	1974
Academic Ability (U1)	0.8	0.9			3.8	3.9	1,1	1,2	68	82
Academic Industry (U2)			0.8	0.9	4.0	4.0	1.2	1.4	72	82
Maturity/Stability (U3)			0.9	0.8	4.2	4.2	1.2	1.2	78	66
Comprehension of Physics (P1)	0.8	0.9			3.6	3.9	1.1	1.2	82	89
Ability to apply (P2)	0.9	0.9			3.5	4.0	1.1	1.1	82	85
Analyze and Evaluate (P3)	0.9	0.8			3.2	3.6	1.2	1.3	84	81
Conduct experimental work (P4)	0.6	0.5	0.5	0.6	3.7	3.9	1.1	1.2	61	64
Process experimental data (P5)	0.7	0.7	0.5	0.5	3.4	3.7	1.2	1.2	78	80
Attitude to physics (P6)	0.5		0.7	0.8	3.9	3.9	1.2	1.4	71	85

Percentage variance accounted for by each factor

Unrotated	63	69	13	10
Rotated (Varimax)	46	45	30	34

At the end of 1973 and 1974 the lecturers involved in the university physics course independently assessed the students on the nine criteria of Table 1. This assessment was made before the students sat their final examinations and without sighting the teachers' earlier subjective assessment of the same students based on the students previous year's study. Lecturers' independent assessments were in good agreement; inter-assessor reliability $r = 0.85$. Where lecturers' individual assessments of a student initially differed, discussion between the lecturers, and with tutors, resulted in a mutually agreed to score being decided upon. A principal components analysis of lecturers' assessments was then carried out for each year separately. As with the teachers' subjective assessments, two significant factors were obtained (unrotated eigen values >1.0). A Varimax rotation of these factors resulted in the factor loadings of Table 3.

TABLE 3

Principal components analyses of lecturer's subjective assessments of first year physics students in 1973 (N = 101) and 1974 (N = 87) with factor loadings ≥0.5

Objective/Criteria	Factor				Mean 0-6 Scale		Standard Deviation		$100h^2$	
	I		II							
	1973	1974	1973	1974	1973	1974	1973	1974	1973	1974
Academic Ability (U1)	0.9	0.9			3.8	3.6	1.4	1.3	83	80
Academic Industry (U2)			0.9	0.9	3.3	3.5	1.6	1.6	87	89
Maturity/Stability (U3)			0.9	0.8	3.5	3.3	1.5	1.2	83	79
Comprehension of Physics (P1)	0.8	0.9			3.4	3.5	1.5	1.4	92	92
Ability to apply (P2)	0.9	0.8		0.5	3.5	3.0	1.5	1.6	91	92
Analyze and Evaluate (P3)	0.9	0.9			3.1	2.9	1.5	1.7	92	91
Conduct exp. work (P4)	0.7	0.6	0.5	0.7	3.4	3.5	1.4	1.4	81	83
Process exp. data (P5)			0.9	0.9	3.3	3.4	1.5	1.7	88	91
Attitude to physics (P6)			0.8	0.8	3.4	3.5	1.4	1.5	82	85

Percentage variance accounted for by each factor

Unrotated	73	73	14	12
Rotated (Varimax)	45	41	39	45

Comparison of Table 2 with Table 3 shows a definite similarity between the teachers' and the lecturers' perceptions except that "objective" P.5 was seen by lecturers as more dependent on "motivation" than it was by teachers. This is to be expected, as less explicit pressure is likely to be placed on a student to complete experimental reports and data analyses at university than at school. Apart from this difference, factor loadings are similar for the four analyses and clearly show that the two factors, "ability" and "motivation", account, in different proportions, for most of the perceived variation among students with respect to all physics "objectives" considered.

The usefulness of the subjective assessments

The predictive usefulness of the teachers' subjective assessments (T) and the lecturers subjective assessments (L) has been estimated by calculating the correlations* between the subjective scores and the students total marks in the first year University physics final examinations (the total marks from two three-hour papers designed to assess the cognitive objectives of the course). The results are shown as MARKS (T) and MARKS (L) respectively in Fig. 2. It must be remembered that teachers' assessments were based on observing students a year prior to the examinations while the lecturer assessments were made on impressions gained during the year of the examinations. The lecturer assessments would therefore be expected to be more highly correlated with students' examination performance than the teacher's assessments. However, for both lecturers and teachers the variation in the correlation coefficients across objectives is strikingly similar. The importance of the comprehension of basic concepts (P.1) is again emphasized.

For the students involved in this study the national school physics examination results (Bursary Physics) correlated 0.80 with their subsequent university physics examination results. When the subjective assessment factor scores "academic ability" and "motivation" were included in a multiple linear regression analysis with the Bursary Physics results, in order to predict the University examination results, a multiple R of 0.81 was obtained. Hence the teachers' subjective factor scores did not significantly increase prediction over and above the school examination result.

Product-moment correlation coefficients are valid only where both variables are normally distributed and the variables are measured on a continuous metric scale. In the results reported above, these conditions have been largely met and where they do not strictly hold the reported coefficients are not considered to be significantly in error. However these conditions were not met for a skewed dichotomous variable labelled COMPLETE which was generated by assigning a mark of 1 to all students (N = 166) who completed all formal components of the university course and a mark of 0 to all students (N = 22) who did not complete all aspects of the course e.g. did not sit the final examination. Results obtained by using this variable must be treated with considerable caution, however the relative differences in correlation coefficients between the teachers subjective assessments and COMPLETE (see Fig. 2 – COMPLETE (T)) are of interest. They give an indication of the relative usefulness of the subjective assessments of different objectives for the purpose of predicting the likely failure of a student to complete the university physics course. In Fig. 2, COMPLETE (T) shows a markedly different profile to MARKS (T). It is interesting to note that the correlation coefficient between Bursary Physics and COMPLETE was 0.26 and between the teachers subjective "motivation factor scores" and COMPLETE was 0.32.

A second variable was generated at the end of the first year at University from students stated intentions for the following year. Scores were assigned to students as follows: those intending to, (i) major in physics = 4, (ii) major in a physics based profession = 3, (iii) major in science without further physics = 2, and (iv) discontinue a science degree = 1. This variable, FUTURE PHYSICS, was approximately normally distributed and not significantly skewed. The correlation coefficient of this variable with the teacher's subjective assessments is given in Fig. 2. FUTURE PHYSICS (T) again shows a significantly different profile from

---

*In this work the mean product moment correlations were obtained from the separate results for 1973 and 1974, by correctly weighting the results for each year in terms of student numbers.

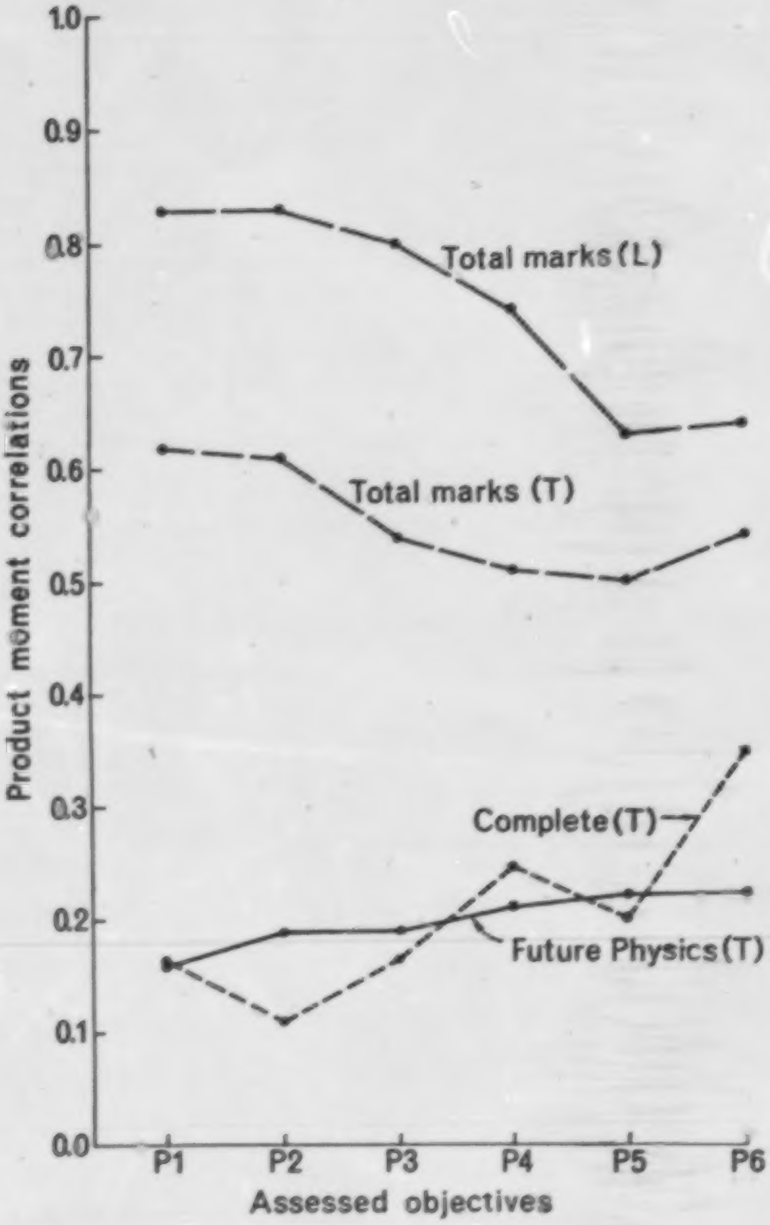


FIGURE 2

Correlations of the subjectively assessed scores with dependent course variables.

the previous results. In attempting to predict FUTURE PHYSICS using multiple linear regression, the teachers' two subjective factor scores were found to be just as inadequate as previous examination performance in Bursary Physics and Bursary Mathematics. In both cases the multiple correlation coefficient was 0.25.

Conclusions

It is clear that in this particular educational system teachers' and lecturers' subjective assessments of each student on the cirteria given, were based primarily on two subjective impressions of the student; the students 'academic ability' and his, or her "motivation for academic work". A major proportion of the variance on the physics subjective scales could be explained in terms of these two factors. However the factor scores did little to increase the prediction of success in University physics examinations over and above prediction based on school physics examination results. There was some indication that subjective assessments may have a potential use in predicting non-examination 'attributes' of importance, for example drop-outs and future career intentions. Undoubtedly much greater reliability and validity of student scores related to objectives needs to be obtained both via examinations and subjective assessments if future studies are to yield really useful student profiles.

References

C.D.U. *Curriculum Development Unit Bulletin Number 48.* Wellington: Government Printer, 1973.

HARLEN, W. *Science 5-13: A formative evaluation.* London: Macmillan Educational, 1975.

KLOPFER, L.E. Evaluation of learning in science. In Bloom, B.S. et al. *Handbook of Formative and Summative Evaluation of student learning.* New York: McGraw-Hill, 1971.

OSBORNE, R.J. & FREYBERG, P.S. The systems approach and its application to curriculum research. *Research in Science Education,* 1977, 7, 71-81.

Research in Science Education, 1977, 7, 165-171.

THE STUDY OF STUDENT MISUNDERSTANDINGS IN THE PHYSICAL SCIENCES

J. K. Gilbert

The, apparently widespread, reduction in the popularity of physical science courses, at both secondary and tertiary levels, has had a number of effects. Attempts by schools to provide a broadly attractive programme in the sciences seem to have reduced the amount of time spent on the physical sciences. Moreover, the introduction of applied and technological aspects, however desirable, within a fixed time allowance, must reduce the emphasis laid on fundamental aspects. Higher education, faced with falling demand, might be forgiven for relaxing entry qualifications and thus taking in students from an ever-widening range of achievement and ability bands. If the standards to be reached by the end of undergraduate courses are to be maintained, then higher education teachers face a steadily more difficult task. Discussion with such teachers soon leads to the comment "undergraduates now just don't understand the basic ideas" and it is to this area that I wish to address myself.

Of course, teachers have always complained about their students, particularly where the two groups had different ends in view. The field may be clarified by seeking out a few definitions. By 'basic ideas' I mean those scientific concepts, e.g. the mole, potential difference, which underpin all advanced study but which are not normally, at least in the U.K., taught in higher education. Students are presumed to have understood them at an earlier stage, but, where this is not justified, the causes for later failure may be found. The meaning of "don't understand" is equally problematic: those teachers who interpret it as a 'failure of knowledge', in Bloom's terms, will use the approaches of the Physics Interface Project (1977) and of Galton et al (1976) to identify and remedy deficiencies. Although 'knowledge' is a necessary component of 'understanding', many teachers will not feel that their students have mastered the 'basic ideas' until an appreciation at Bloom's higher order level has been achieved.

In looking for the sources of student misunderstanding, the first question becomes: are we seeing the problem of 'no understanding' which stems from a simple lack of teaching at an earlier stage? If this is the case, the remedy lies in a primary teaching approach, so that a coherent scheme, such as one based on Gagné (1970) may be appropriate. However, if it is 'faulty understanding', resulting from incorrect teaching and learning at an earlier stage, that has become apparent, diagnosis must precede re-education. Higher education teachers, working on evidence that is seldom revealed, favour this latter interpretation, but infrequently follow it to the conclusion suggested above. The fragile ground of collecting evidence, or at least, publishing it, has only been lightly walked upon. Warren (1976), working with a small group of trainee secondary school physics teachers, found that about one-half misunderstood the mass-energy relation.

Other potential sources of misunderstanding readily present themselves. The use of "problems" in both text-books and examinations presents ample scope for the promotion of error: Warren (1971a) states that "many text-book calculations are physically improbable or impossible: the presumption of long-obsolete standards of instrumentation, the use of inappropriate scales of numerical size for the problems, coupled to an over-simplification of theory, leads to a distortion of reality". Furthermore, the whole area of language use in science and science education is full of pitfalls for the unwary (Gardner, 1974).

One explanation of student misunderstandings may lie in the area of intellectual development. Piaget's work is of central importance here and it is to be regretted that he has done relatively little work with older adolescents and adults. Only a few studies have been undertaken on higher education students, e.g. Elkind (1962), McKinnon and Renner (1971), Renner and Lawson (1973), Towler and Wheatley (1971), and most of those in American Institutions where access policies and the cultural background make transfer of conclusions to other situations difficult. The main question concerns the distribution of capacity for 'formal operational' thought as opposed to the subordinate 'concrete operational' variety. Arons (1976) gives representative results: "not more than 25% of the students (in the sample) have attained the level of formal operations, perhaps 25% are in transition between the concrete and formal levels, and about 50% are essentially concrete operational".

The implication must be that, if the intellectual demand level of concepts, when appreciated in the upper reaches of Bloom's taxonomy, exceeds the developed intellectual capacity of the student, then effective learning cannot, and has not, taken place. The problem of the diagnosis of misunderstandings can thus be seen to have three components: finding the 'Piagetian level' of the student, estimating the 'demand level' of concepts at the higher orders of appreciation, and investigating the consequences for students of earlier 'intellectual stress' in relation to particular basic concepts.

The research tools used by the Geneva School are too well documented to warrant rehearsal here, but Karplus et al (1977) have produced a particularly readable interpretation. The estimation of the demand level of a concept should be helped by the procedure of concept mapping, that is by producing a network of hierarchically related concepts, for which several schemes exist, e.g. Klausmeier et al (1974), Lewis (1974), Markle and Tiemann (1970). Estimation of the demand level for various scientific concepts has been attempted by Malpas and Brown (1974) for mathematics, and by Shayer for the Nuffield schemes in chemistry (1973), physics (1972) and biology (1974) respectively: they all depended on expert judgements and the criteria used were not published.

The Institute for Educational Technology at the University of Surrey has been interested in these matters for some time. A new phase of work, done jointly with Visiting Staff, began a year ago and centred initially on the third component of diagnosis: finding out what misconceptions students had acquired on admission to undergraduate courses. This tentative work, of which a very preliminary account is given below, had a number of aims: to investigate the usefulness of different question types for diagnosis, to look for patterns of performance within higher education and for trends between school and higher education, to explore types of result anlaysis and presentation likely to be of greatest use to practising teachers, and to seek a formula for future diagnostic procedures.

The concepts chosen, after discussion with practising teachers, were 'the mole' in chemistry and 'potential difference' in physics. Collaborating groups worked on each concept, so that similarities and differences are reflected in the work. The preliminary preparation for the construction of the tests rested basically on the extensive teaching experience of those involved. However, in the case of 'the mole' a chart of all those situations in which the concept is encountered was drawn up, and in the case of 'potential difference' a concept map was prepared. It was decided to use those question types employed in the Advanced Level examination, because the concepts had, most likely, not been taught in higher education, and because the basic framework of examples was thus readily available.

Open-ended questions have been used with undergraduates by Warren to investigate circular motion (1971b) and heat (1972), whilst Mannino et al (1973) looked at 'electric interactions' with Italian school children. The attraction of the approach is that students have an opportunity to indicate how they arrive at conclusions, but the difficulties stem from attempts to classify these processes. In the event, one such question was included in the 'potential difference' test. Multiple-choice tests, with specially constructed distractors, have been used by Driscoll (1975) to investigate chemical equilibrium with Australian undergraduates, by Duncan and Johnstone (1973) to look at 'the mole' with Scottish school children and by Za'rour (1975) for a broad study of high school and university students in the Lebanon. In the tests, the majority of questions were of the 4- and 5- alternative type, with some of the assertion-reason type.

Attempts were made to administer the two tests to identical groups of students, but this was not always possible. Within higher education a broad variety of class types were used, with some emphasis on applied and technological orientation. Sixth Form classes from a wide variety of schools were included. All the students involved were 'volunteers', having been told the object of the exercise. No time limit was placed on either test nor were circumstances dictated, in view of the wide range of situations in which operational diagnostic tests might be taken. In practice, 'the mole' was taken by 212 students (137 from higher education and 75 from schools) and the 'potential difference' test by 138 students (87 from higher education and 51 from schools). All students did the tests on their own and most in private study time, although some school students had a time constraint imposed. A summary of the overall test data is given in Table 1.

The 'Mean Advanced Level' score is the mean value, for the higher education class, of attainment in the most relevant subject, e.g. physics for university physics courses, using the scale Grade A–E becoming Value 5–1. The first conclusion to be drawn concerns the low mean achievement that is apparent. The problems of higher education teachers become even more evident when the range within a class is considered: thus the Industrial Chemists at University A have a mean value of 1.9 and a standard deviation of 1.6, the total number of students being 35! The mean scores for 'the mole' test seem fairly satisfactory: one would hardly expect University Physicists to do well on it as the subject is no longer of everyday relevance. Preliminary examination of the score distribution shows that the school-level classes are narrowly clumped, although the focus varies from school to school. The scores of individuals in higher education classes show a very much broader spread: again the teachers' problems become evident. The scores for the 'potential difference' test, which was conceived as being the more searching of the two, are not satisfactory and the understanding of this concept, which is basic to all the physical sciences, is seen as inadequate when demands beyond the algorithm-application stage are made. However, it is the poor overall performance of the 'College of Education E' students which seems the most serious outcome: these students were in the fourth (and honours degree) year of a programme intended to train secondary school science teachers. The pattern of scores from the schools does, however, seem relatively satisfactory when compared to the poor level of improvement evident in higher education: the lack of post-school teaching in these concepts does seem to have been partially substantiated.

In discussing the overall analysis of the tests, remarks will be restricted to 'the mole' as 'potential difference' is rightly the province of its originators: Dr. H. Helm of Rhodes University, South Africa, and Dr. B. McInnes, Sydney University. In analysing data from 'the mole' test, facility indexes were calculated, in respect of all higher education institutions and all schools, for each question. The grand mean facility indexes led to the grouping of questions as follows: 100-75% (seen as representing mastery), 74-50% (seen as representing a roughly adequate performance), 49-25% (seen as an inadequate performance), 24-0% (seen as quite

TABLE 1

Summary of performance on the diagnostic tests

Institution	Nature of Class	N (Max.)	Mean Age (Yrs.)	Mean 'A' Level Score	The Mole		Potential Difference	
					Mean Time (Mins.)	Mean Score (Max. 34)	Mean Time (Mins.)	Mean Score (Max. 31)
(i) *Higher Ed.*								
University A	Industrial Chemistry	35	19.7	2.0	52	25.8	40	11.8
University A	Physics	30	19.3	0.5	32	17.1	45	15.6
University B	Chemistry & Chem. Engineering	29	19.4	2.2	51	28.3	-	-
University B	Metallurgy	25	18.4	1.7	52	26.1	-	-
University C	Physics	11	18.1	1.4	-	-	51	17.1
University D	Chemistry	3	19.3	1.3	-	-	-	15.7
Coll. of Education E	Trainee Teachers	18	21.6	1.5	50	23.0	46	15.6
(ii) *Secondary*								
M. 11 - 18 Comp.	Chemistry	11	16.9	-	27	24.2	59	13.9
N. 12 - 18 Comp.	Chemistry	13	17.4	-	47	25.9	48	13.8
P. 16 - 19 Comp.	Chemistry	25	17.7	-	33	19.3	33	13.6
Q. 13 - 18 Ind.	Chemistry	26	17.1	-	44	27.4	-	-

inadequate). Within each group the most commonly selected distractor was identified for each question, and the question examined further if more than 10% of the students selected this distractor. The arguments surrounding random choice from these five alternative questions were disregarded, in view of the nature of the test, the students and the administrative arrangements.

Fifteen questions (out of the whole test of 34) fell into the 100-75% Facility Index Group, and only four questions in this group of 15 had distractors that were frequently chosen. These four questions covered: the application of Avogadro's hypothesis, and variations on the theme of concentration in solutions. This latter group may stem from the habit in schools of using ready-prepared solutions in volumetric work: remedial action might involve the older tradition of the preparation of accurate solutions from solid materials.

The second, or 74-50% facility, group contained 14 questions, of which 11 had single distractors of over 10% popularity. Although covering a range of situations, three kinds stand out: firstly, the calculation of mole quantities for liquid materials, secondly calculations involving the use of equations for solids, liquids, and gases, and thirdly a simplistic definition of the mole itself. Although the definitional problem may not be seen as, in reality, of practical importance, those calculations involving reaction quantities are extremely important and might be rectified by practical stoichiometric reaction exercises.

Three questions only came within the 49-25% group and the distractors, if chosen by an alarmingly high proportion of students, represent situations where the problem represents a sequence of assumptions and calculations placed end-to-end. Only two questions fall within the 24-0% facility level, but one of them is of concern: a majority of students do not know the definition of *molarity* and, evidently because of a failure to appreciate that solutes have a finite volume, chose the distractor corresponding to *molality*.

Looking at the mole test overall, the most interesting thing was the absence of general improvement between schools and higher education: school students did 23 of the questions more correctly as compared with higher education students (the mean facility index gap being 8% points), whilst only 11 questions constituted the converse group (the facility index gap being 7% points). Evidently it is not taught and the above data indicates that it might well be, perhaps by practical work related to the operations of the analytical chemist.

These preliminary tests have indicated some pointers for future work in the same area. The use of open-ended questions to generate important distractors for subsequent multiple-choice tests seems called for, as does the inclusion of space for students to explain their working. In the case of 'the mole', it would seem sensible to break down the test into sections which could be applied as relevant, e.g. metallurgists are concerned with solids, and only infrequently with solutions. On the other hand, there are attractions to multi-concept tests, with questions on one area arranged in order of increasing demand, for application to new entrants.

On the broader, and more theoretical, front a number of indications have emerged for work shortly to begin at the University of Surrey. Piagetian testing of freshmen should be interesting, even if the results will have to be handled carefully. Studies of the 'demand level' of a particular concept, and the Piagetian level of the student, similar to the work of Lawson and Renner (1975), should be rewarding, particularly if they can be applied to students as they move from school to higher education, in a logitudinal study. It should also be possible to build up a picture of the concept map internalised by a number of students, by the use of a statistical approach to a multi-concept test.

There must remain little space here to emphasise that all such work must have, as its goal, the remedy of such problems as do emerge. In the current literature are to be found a number of remedial approaches, ranging 'how to fill knowledge gaps', e.g. A.C.S. Symposium (1976), through 'courses with a Piagetian flavour', e.g. Bass, Dever and Diboll (1976), to 'conceptual level promotion', e.g. Lawson and Wollman (1976). The evolution of multi-faceted remedial approaches, perhaps involving the use of individualised learning packs allied to tutorials, open access laboratories, and computer simulations, will be demanding but rewarding. Their use will, in addition to the obvious purpose, offer opportunities for further 'research in the context of development'.

References

A.C.S. *Symposium on teaching chemistry to underprepared students.* San Francisco, 1976.

ARONS, A.B. Cultivating the capacity for formal reasoning: objectives and procedures in an introductory physical science course. *American Journal of Physics*, 1976, **44**, 834-838.

BASS, W.T., DEVER, D.F. & DIBOLL, A.G. A remedial programme and Piagetian certification of its success. *A.C.S. symposium on teaching chemistry to underprepared students.* San Francisco, 1976.

DRISCOLL, D. *Science Teacher Education Project*. London: McGraw-Hill, 1975. Case Study C, Unit TNK 4.

DUNCAN, I.M. & JOHNSTONE, A.H. The mole concept. *Education in Chemistry*, 1974, **10**(*6*), 213-214.

ELKIND, D. Quality conceptions in college students. *Journal of Social Psychology*, 1962, **57**, 459-465.

GAGNÉ, R.M. *The conditions of learning.* (Second edn.). New York: Holt Rinehart and Winston, 1970.

GALTON, M.J., HOLLOWAY, J.H., RAYNOR, J.B. & TOMLINSON, M.J. Introductory chemistry at university. *Education in Chemistry*, 1976, **13**(*2*), 38-40.

GARDNER, P.L. Language difficulties of science students. *Australian Science Teachers' Journal*, 1974, **20**(*1*), 63-76.

KARPLUS, R., LAWSON, A.E., WOLLMAN, W., APPEL, M., BERNOFF, R., HOWE, A., RUSCH, J.J. & SULLIVAN, F. *Science teaching and the development of reasoning.* Berkeley: University of California, 1977.

KLAUSMEIER, H.S., CHATALA, E.S. & FRAYER, D.A. *Concept learning and development: a cognitive view.* New York: Academic Press, 1974.

LAWSON, A.E. & RENNER, J.W. Relationship of science subject-matter and developmental levels of learners. *Journal of Research in Science Teaching*, 1975, **12**(*4*), 347-358.

LAWSON, A.E. & WOLLMAN, W.T. Encouraging the transition from concrete to formal cognitive functioning – an experiment. *Journal of Research in Science Teaching*, 1976, **13**(*5*), 413-430.

LEWIS, B.N. *New methods of assessment and stronger methods of curriculum design*. I.E.T.: Open University, July 1974.

MALPAS, A.J. & BROWN, M. Cognitive demand and difficulty of G.C.E. O-level maths pretest items. *British Journal of Educational Psychology*, 1974, **44**, 155.

MANNINO, S., PALMA-VITTORELLI, M.B. & SPERANDEO-MINEO, R.M. An investigation on the understanding of concepts related to electric interactions at High School level. *Seminar on the Teaching of Physics in Schools*, 1973, **2**, 102-114.

MARKLE, S.M. & TIEMANN, P.W. *Really understanding concepts: or in frumious pursuit of the jabberwoch*. Champaign, Illinois: Stipes, 1970.

McKINNON, J.W. & RENNER, J.W. Are colleges concerned with intellectual development? *American Journal of Physics*, 1971, **39**(*9*), 1047-1052.

Physics Interface Project. Cardiff: University College Cardiff Press, 1977.

RENNER, J.W. & LAWSON, A.E. Promoting intellectual development through science teaching. *The Physics Teacher*, 1973, **11**, 273-276.

SHAYER, M. Conceptual demands in the Nuffield O-level physics course. *School Science Review*, 1972, **51**, 26.

SHAYER, M. Conceptual demands in Nuffield O-level chemistry. *Education in Chemistry*, 1973, **8**, 8.

SHAYER, M. Conceptual demands in the Nuffield O-level biology course. *School Science Review*, 1974, **56**, 381.

TOWLER, J.O. & WHEATLEY, G. Conservation concepts in college students: a replication and critique. *Journal of Genetic Psychology*, 1971, **118**, 265-270.

WARREN, J.W. Thinking quantitatively. *Physics Education*, 1971a, **6**, 238-241.

WARREN, J.W. Circular motion. *Physics Education*, 1971b, **6**, 74-77.

WARREN, J.W. The teaching of the concept of heat. *Physics Education*, 1972, **7**, 41-44.

WARREN, J.W. The mystery of mass-energy. *Physics Education*, 1976, **11**(*1*), 52-54.

ZA'ROUR, G.I. Science misconceptions amongst certain groups of students in Lebanon. *Journal of Research in Science Teaching*, 1975, **12**(**4**), 385-391.

Research in Science Education, 1977, 7, 173-180.

THE AIMS OF FIRST YEAR CHEMISTRY COURSES, THE EXPECTATIONS OF NEW STUDENTS AND SUBSEQUENT COURSE INFLUENCES

P. P. Lynch and G. C. Gerrans

Introduction

Previous work (Frazer, Palmer & Sleet, 1976) examined the attitudes of undergraduates towards their courses by investigating the degree of discordance between their rating of the *observed importance* of specified aims with their rating of the *preferred importance* of the same aims. For all but two of the thirty-three aims listed the scores for preferred importance were significantly greater than the scores for observed importance. Thus, according to the students, there was inadequate opportunity to develop intellectual skills, problem solving abilities etc., and undue emphasis placed upon the acquisition of knowledge and the degree as a form of certification.

The measuring instrument (which used a six point Likert scale) did not directly allow for priorities between aims to be stated. Individual students' rank ordering of the thirty-three aims was not, however, considered. Priorities within these aims could have been deduced indirectly by comparing inter-aim means. However, on completing such an examination later, several unexpected results appeared. For example, it would seem that the students felt there was inadequate opportunity for developing the skills associated with working safely in the laboratory and that these skills were also important. Experience would indicate that, whilst this is laudable, it is seldom considered as a priority task when weighed against the acquisition of other laboratory skills such as the ability to make careful observations and the ability to carry out manipulative skills. It seemed therefore that the use of the Likert scales, in conjunction with the instruction to give an aim its preferred weighting, resulted in indications of what the students thought to be *ideally desirable* and not what was *pragmatic* within the context of a degree course of finite length. In order to investigate the latter considerations further we decided to re-examine staff/student interaction in relation to the following questions.

How do students about to start their university careers perceive their intended courses? What do they expect of them and how do these expectations match the views of their future teachers?

In an attempt to gain an insight into the views of new (beginning) students and their teachers within the context of having to draw priorities for the learning task, a pilot study was undertaken at the University of The Witwatersrand.

In the present extension the list of aims was rewritten and arranged into four groups:

Group A : aims associated with the acquisition of knowledge
Group B : aims associated with the acquisition of cognitive skills
Group C : aims associated with laboratory work
Group D : aims related to students' attitudes

Questionnaire to staff

The list of 32 aims (which are shown in Table 1) were arranged randomly within the

TABLE 1

The questionnaire on aims given to lecturing staff

Aims related to knowledge

Select the FOUR most important aims from the list below. Place a tick in the appropriate boxes to indicate your choice.

1. To improve knowledge and understanding of the historical development of chemistry.
2. To improve understanding of the basic principles of chemistry.
3. To increase awareness of the economic implications of chemistry.
4. To show how chemistry is related to other science subjects.
5. To improve understanding of the application of chemistry to industry.
6. To increase factual knowledge of chemistry.
7. To increase awareness of the moral and social implications of chemistry.
8. To improve understanding of the application of chemistry to environmental problems.
9. To improve knowledge of the organisation and functions of the chemical industry.
10. To improve understanding of how theories can explain experimental observations.

Aims related to cognitive skills

Select the FOUR most important aims from the list below. Place a tick in the appropriate boxes to indicate your choice.

11. To improve the ability to think in abstract terms.
12. To improve the ability to think logically.
13. To improve the ability to learn independently.
14. To improve the ability to use new approaches or ideas when called upon to solve a problem.
15. To improve skills in oral communication.
16. To improve the ability to solve new problems in chemistry by using basic principles and concepts.
17. To improve skills in written communication.
18. To develop the ability to design experiments.
19. To teach how to obtain relevant information from the literature and other sources.

Aims related to experimental skills

Select the THREE most important aims from the list below. Place a tick in the appropriate boxes to indicate your choice.

20. To improve the ability to work safely in a laboratory.
21. To improve manipulative skills in the laboratory.
22. To improve the ability to make careful observations of experimental phenomena.
23. To improve the ability to conduct laboratory experiments in an organised way.
24. To improve the ability to interpret experimental observations in a logical way.
25. To increase awareness of the importance in chemistry of accuracy in experimental observations.
26. To increase awareness of the importance of writing an accurate laboratory report.

Aims related to student attitudes

Select the THREE most important aims from the list below. Place a tick in the appropriate boxes to indicate your choice.

27. To develop an ability to undertake employment in many fields of chemistry (or related to chemistry).
28. To develop an ability to do research or original work.
29. To develop an ability to work with a team or in a group.
30. To increase interest in the theoretical and conceptual structure of chemistry.
31. To develop open-mindedness about careers outside chemistry.
32. To increase interest in experimental work.

four distinct groups. In Group A there were ten aims, in Group B nine aims, in Group C seven aims and in Group D six aims. Each section was to be considered separately. Staff were asked to indicate the four most important aims in Group A, the four most important aims in Group B, the three most important aims in Group C and the three most important aims in Group D. In each group there was the opportunity for the staff member to indicate that he had no particular preference amonst the aims.

The response rate for the faculty was sixteen out of twenty-five.

Questionnaires to students

The aims were modified to that of the point of view of the learner to produce two further questionnaires intended to measure the course *expectations* and its subsequent *influence*. Each aim is associated with a corresponding expectation and a corresponding influence e.g.

Staff : Aim 6 – To increase factual knowledge of chemistry
Student : Expectation 6 – To increase my factual knowledge of chemistry
Student : Influence 6 – Increased my factual knowledge of chemistry
Staff : Aim 28 – To develop an ability to do research or original work
Student : Expectation 28 – To develop, in me, an ability to do research or original work
Student : Influence 28 – Developed my ability to do research or original work.

The purpose of the questionnaire, given in its introduction, was read to the students during a short preamble aimed at winning their cooperation. Students were asked to select the *most important* influences from any particular group according to the instructions. So for instance in the case of 'influences related to knowledge' students selected four items from a list of ten. Students were taken through the questionnaire section by section and asked to read each set of items carefully and several times before making any selection. In this way the completion of the questionnaire, excluding the introductory preamble, took approximately 20 minutes during a normal, but previously undisclosed, lecture period.

A sample of students was also interviewed to check, amongst other things, their interpretation of the work *"expectations"*. They understood it to mean, *"What I hoped for"*. This was in accordance with our intentions. In addition it became clear from the interviews that students' perception of tertiary level work was based mainly on their experience of secondary science and was also affected by the views of teachers and the media etc.

The sample

Replies were obtained from 136 newly enrolled students (response rate 100%) from the chemistry major class at the University of the Witwatersrand at the beginning of the 1976 session.

Replies to the 'influences' questionnaire were obtained from 109 first year students from the chemistry major class, at the end of the 1976 session (response rate 80%, based on the number of students writing the final examination). This particular group was made up of science/industrial chemistry students (59), and chemical engineering/metallurgy/mining geology students (50) pursuing the same first year course. The major course is the normal route by which students proceed towards a degree in chemistry, industrial chemistry or chemical engineering.

Analysis and display of results

The forced choice technique employed for the questionnaires allows the relative importance of the aims, expectations and influences to be established from the ratio

$$\frac{n}{N} = \frac{\text{Number of respondents choosing the item}}{\text{Total number of respondents}}$$

The $\frac{n}{N}$ values for aims of staff versus course influences on students, aims of staff versus expectations of students and expectations of students versus influences on students are plotted as scattergrams (Figs. 1, 2 and 3 respectively). A measure called the discordance, defined as the difference between, say $\frac{n}{N}$ (aims of staff) and $\frac{n}{N}$ (influences on students) was calculated in all cases. The discordance values for almost all items in Figs. 1, 2 and 3 are greater than those observed between the different student groups. For that reason the students have been treated as a homogeneous group. But it is interesting to note that the two student groups (science/industrial chemistry and chemical engineering/metallurgy/mining geology) show relatively little difference in their expectations of the course or in their perceptions of the influence of the course within the framework of the items listed in the questionnaires.

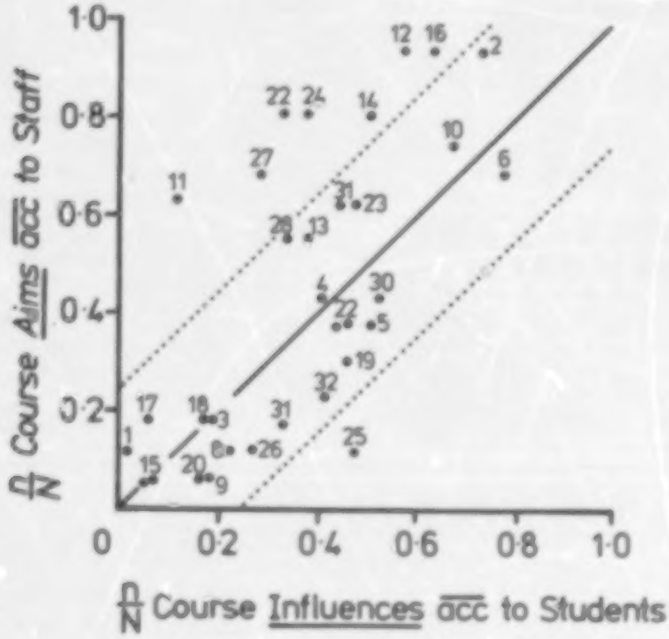


FIGURE 1

Scattergram for $\frac{n}{N}$ (Student Jnfluences) vs $\frac{n}{N}$ (Staff Aims)

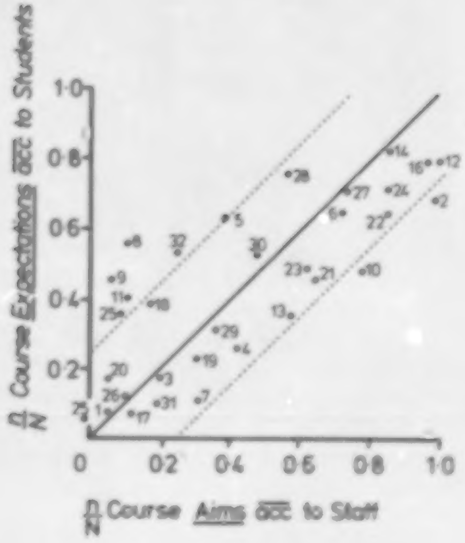


FIGURE 2

Scattergram for $\frac{n}{N}$ (Staff Aims) vs $\frac{n}{N}$ (Student Expectations)

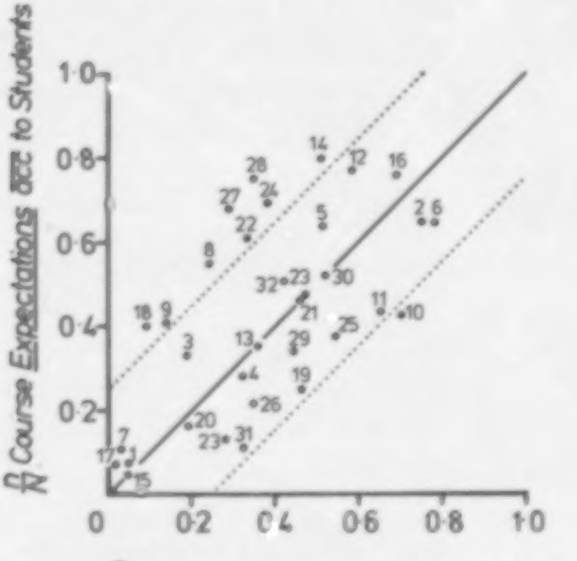


FIGURE 3

Scattergram for $\frac{n}{N}$ (Student Influences) vs $\frac{n}{N}$ (Student Expectations)

Discussion

These results afford three interesting comparisons namely:

1. Course influences versus staff aims (Fig. 1)
2. Student expectations versus staff aims (Fig. 2)
3. Student expectations versus course influences (Fig. 3)

These comparisons represent important interactions between the lecturer and student and are of considerable relevance when course design or course revision is being considered.

In order to facilitate discussion we have concentrated mainly on Fig. 1 (influences versus aims) thus focussing attention on the *direct* interaction between lecturer and student. Fig. 2 and 3 though subsidiary in this respect are nevertheless important. Fig. 2 (expectations versus aims) tells us something about the students perception of tertiary level chemistry as a consequence of school work, media etc. Fig. 3 (expectations versus influences) gives a measure of the change of emphasis as perceived by the student.

The data afforded by each scattergram can, with advantage, be discussed under three separate headings, given below.

Important aims, influences or expectations associated with small discordances

When $\frac{n}{N}$ values for both coordinates involved are about equal and greater than 0.6 then this indicates items which are considered to be important and about which there is concordance.[1] In Fig. 1, this corresponds to items 2, 6 and 10, in Fig. 2 to items 2, 6, 12 and 16. An interpretation of Fig. 1 would be that staff consider aims such as improving the understanding of basic principles, increasing factual knowledge and improving understanding of the relationship between theories and experimental observations to be of major importance. Student opinion indicates that these do in fact correspond to major influences as far as the course was concerned.

Important aims, influences or expectations associated with large discordances

The dotted lines in the scattergram mark the boundaries of a discordance > 0.25. This is an arbitrary boundary beyond which differences are greater than 25% between the different coordinates.[2] There are 7 aims of staff (11, 12, 14, 16, 22, 24 and 27) which exceed the corresponding course influences according to students. It is interesting to note that these discordances are not associated with knowledge but with 'cognition', 'experimental work' and 'attitudes'. There are considerable differences between staff aims related to cognitive skills and students' perception of the magnitude of the corresponding course influences. Four out of the nine items (items 11-19 inclusive) are markedly different.

The discrepancies associated with practical work concern items 22 and 24. Staff feel that 'the ability to make careful observations' and 'to interpret observations in a logical way' are of major importance – in fact the two most important aims as far as practical work is concerned. Students indicate the corresponding influences are much lower and associated with some of the greatest discordances observed for any items. One of the authors has had cause to comment on this effect in some detail elsewhere (Lynch, 1977). The discordance observed for item 27, 'developed my ability to undertake employment in many fields of chemistry (or related to chemistry)', merits further investigation in the light of present employment prospects.

Its implications are not immediately obvious. It is interesting to note that staff consider this to be the most important aim within the group of aims related to attitudes, whereas students do not consider it to have been a major influence.

Aims, influences or expectations considered to be relatively unimportant by both staff and students

There are a number of items which fall into this category. It is interesting to note that except for aims 11 and 25, there is considerable concordance between staff and students (Fig. 1) on these relatively unimportant aims/influences. The question is whether some of these items should be considered so unimportant in a first year course. Should the historical, economic, social and environmental aspects of chemical knowledge receive low priority at this stage? (items 1, 3, 7 and 8) Is knowledge of the organisation and function of the chemical industry so unimportant? (item 9). Certainly students had much higher expectations of these items. Are written and communicative skills so peripheral as far as first year work is concerned? (items 17 and 18) They would appear to be so.

Conclusion

We feel that these results, particularly when represented in scattergram form, should be of value to us and to our colleagues concerned with first year work. Although the results offer no conclusions for action they should help heighten the level of debate concerning course construction and design by providing some idea of the relative importance of the different aims and by spotlighting unusual features (such as discordances) which may represent real causes for concern. One might be tempted to think that the results of Fig. 2 (aims versus influences) is all that is needed but we feel that one should also allow for, or at least be aware of, the expectations of students. After all science and chemistry teaching begin in the school room not in the university lecture theatre. Knowledge, cognitive skills and attitudes are partially developed prior to our own course work. We recognize that in the final analysis most course construction is based on intuition and is likely to remain so. We are not advocating a systematic behavioural approach but simply suggesting a methodology for obtaining information which should guide and supplement intuition.

Footnotes

1. Primary data in addition to that given in Figures 1, 2 and 3 is available on request from P.P. Lynch.
2. It will be appreciated that 0.6 is an arbitrary value though there is some statistical justification for a value of this order.
   The calcuation of statistically significant discordances is discussed elsewhere (Lynch, P.P., & Gerrans, G.C. *S.A. Journal of Science*, 1977, 73, 25).

References

FRAZER, M.J., PALMER, C.R., & SLEET, R.J. Aims of first degree courses: student opinion. *Education in Chemistry*, 1976, 13, 2.

LYNCH, P.P. The nature, purpose and organisation of practical work in high schools and universities in South Africa. Pretoria: Report of the Human Sciences Research Council of South Africa

Research in Science Education, 1977, 7, 181-191

SOCIAL CHARACTERISTICS, MOTIVATIONS, ACHIEVEMENTS AND CHOICES OF HSC SUBJECTS OF ENVIRONMENTAL STUDIES IN A TERTIARY INSTITUTION – SOME PRELIMINARY FINDINGS

C. B. Kings

Introduction

The results reported in this paper are from the 1976 cohort of students at Rusden State College, with special reference to environmental students. Rusden SCV is one of the eight State Colleges of Victoria which are almost solely concerned with teacher education. Rusden SCV has some 2000 students enrolled in preservice courses, and an increasing number of postgraduate students. The Environmental Studies Department is one of nine departments in which students can enrol. The other departments are Business Studies, Drama, Language and Literature, Mathematics, Media Studies, Physical Education, Social Studies and Home Economics. The data represent part of the total pool of information which has arisen from the Rusden Tertiary Entrance Project (RTEP). This project was set up to investigate alternative means of assessment which might be used for purposes of admitting students to Rusden State College if and when the Higher School Certificate examinations in Victoria are abolished. More specifically, the primary aim of the project is to find out the extent to which scores on the Australian Aptitude Test and the Higher School Certificate predict the first year performance of students in the College. A secondary aim of the project is to find out the background characteristics of students in the College who want to teach. The project is longitudinal and it is hoped to monitor the 1976 cohort for two or three years after students have reached the work force interface. At present the first year results of the 1976 cohort are being analyzed. The present study is to be expanded to include data on the 1977 first year cohort at Rusden State College.

Instruments and procedures

Data were collected from the Vice Chancellor's Questionnaire 'A Study of Students Entering Tertiary Education'. The questionnaire responses provided information about such background characteristics as sex, age, area of student's residence, parents' education and parents' occupation. Responses were also collected from a questionnaire developed at Rusden State College called 'Motivations for Entering a Teacher Preparation Course 1976'. The Australian Scholastic Aptitude Test (ASAT, Series B), a two hour content-free test developed by the Australian Council for Educational Research, was administered in the first week of Term One 1976 and the scores recorded.

The 'Anderson Scores', the scores for individual Higher School Certificate (HSC) subjects, were recorded from the computer print out of the Victorian Universities Admissions Committee (VUAC). The 'Anderson Score' is the aggregate of the applicant's best four subjects at HSC plus 10 per cent of each other subject in which 40 per cent or more was gained. The College first year results of students have also been recorded.

In this paper some of the responses and results of the 71 environmental studies students are discussed and compared with those of all students in the 1976 first year cohort. The total cohort comprised 495 students. It follows that some of the data reported may reflect characteristics of teacher education students in general, whilst others are more specific to

Environmental Studies students.

Sex and age

The Environmental Studies Department had 52 (73 percent) females and 19 (27 percent) males in the original intake. The female bias in the enrolment of this department contributes to the female sex bias in the College (66 percent female). When students of home economics are omitted from the 1976 enrolment, 57 percent of the intake is female. Elsworth (1976) reports that 65 percent of students enrolling in a similar course, Environmental Science, at Melbourne State College in 1975 were female.

Presumably, then, more females are attracted by an environmental studies/science 'type' course. The factors which influence students' decisions in this case are unclear. However, it is also true that more female students have studied biology at HSC (White, 1973), and perhaps similar factors influenced their choice of college course.

Of the environmental studies students, 80 percent (compared with 82 percent for the whole cohort) were under the age of 20 and corresponded to the 80 percent who had qualified in the year previous to enrolment at Rusden. This pattern of studying at the tertiary level directly after leaving school was similar to that for the rest of the cohort (85 percent). This raises a question: Is it advisable that only students who have just left school should be encouraged to apply to Rusden, or should this invitation be more widely extended to mature students? Research in which the College results of 'mature' and 'normal' age students have been compared, have indicated that 'mature' students generally perform better with fewer dropouts (Kings et al., 1976; see also van Helden, 1975; Walker, 1975; Barrett, 1976; Marsh, 1976).

Institution and course preferences

Of the first year environmental studies students 78.6 percent stated that Rusden was the institution of their first choice and 67.0 percent stated that environmental studies was a course of their first choice, compared with 77.0 (institution) and 81.4 (course) percent respectively for the whole College (see Table 1).

At Melbourne State College, Elsworth (1976) found that 45.5 percent of first year students enrolled in environmental science at the College as the course of their first choice. This figure seems to be somewhat lower than Rusden's. Some second year students at Melbourne State College (*op. cit.*) indicated that lack of appropriate prerequisites for a desired course was an important reason for their choice of environmental science.

For Rusden students, the course chosen corresponded to the preferred institution. The only other institution in Melbourne offering a similar course is Melbourne State College, to which some of the other preferences presumably went. The results show that the environmental studies program at Rusden attracted a large number of first preferences. This is in contrast to the university Departments of Physics and Chemistry who often find it hard to fill their quotas. An important reason for the comparative attractiveness of the College courses is that there are no specific subjects required of students who enrol in the courses.

Type and location of school

About half of the environmental studies students; 51.6 percent;[1] attended State

TABLE 1

Number of students who selected Rusden by institution and course as first choice by department

Department	Rusden First Choice		Course First Choice	
	N	Percent	N	Percent
Business Studies	23	63.9	15	41.7
Drama	23	82.1	26	92.9
Environmental Studies	55	78.6	57	81.4
Language & Literature	55	77.5	48	67.6
Maths	10	76.9	10	76.9
Media Studies	26	81.3	27	84.4
Physical Education	75	93.8	77	96.3
Social Studies	18	62.1	77	58.6
Home Economics	69	68.3	67	67.0
Total	354	77.0	344	74.9

High schools, 19.4 percent Roman Catholic schools, and 24.2 percent Independent schools. This departmental patterns differs from that of the total cohort where a higher percentage, 63.0, attended State High schools, 21.1 Roman Catholic schools and 17.1 percent Independent schools.

The percentage of environmental studies students who attended State High schools corresponded more closely to the percentage in universities.[2] It is difficult to explain the differences in students' school background between departments of the College, however, the figures for the whole cohort indicate that State Colleges have a different student population (see also Barnard, 1976). It is of interest to speculate as to how school curriculum offerings, amongst other factors, limit choice of tertiary courses.

In the Environmental Studies Department 60.3 percent of students, compared with 52.7 percent of students for the whole cohort, came from schools in Melbourne.[3] The location of the school and presumably the family, may in some way have influenced the choice of tertiary course for some students.

Sixty-two percent of environmental studies students live at home (with parents or relatives), which would be explained by the fact that most live in Melbourne. There is evidence that some environmental studies students who were offered a place at Rusden and who live outside the metropolitan area, have postponed coming to Rusden, mainly for economic reasons. This would also be true of the whole cohort.[4]

Socio-economic status

Of the parents of environmental studies students, 31.4 percent of fathers and 15.7 percent of mothers have upper or lower professional occupations. See Tables 2 and 3 below:

TABLE 2

Father's occupation

	Upper/lower professional	Large scale/ small scale employer/manager	Non-manual clerical foreman	Semi-skilled, farm workers home duties
	%	%	%	%
Environmental Studies	31.4	22.8	24.2	21.4
Whole Cohort	23.0	26.9	30.4	20.1

TABLE 3

Mother's occupation

	Upper/lower professional	Large scale/ small scale employer/manager	Non-manual clerical foreman	Semi-skilled, farm workers, home duties
	%	%	%	%
Environmental Studies	15.7	4.3	21.4	58.6
Whole Cohort	12.0	4.3	22.5	61.3

The occupations of students' parents show a similar pattern for the environmental studies group and the whole cohort. Anderson and Western (1970, p.9) in comparing the occupation categories of parents of students in engineering, law, medicine and teaching at six Australian Universities found:

> Among the broad occupational categories teacher trainees have the smallest proportion of fathers who are professionals, employers, or managers; the largest who are manual workers. The differentiation is greatest at the 'lower' end of the distribution: 19 percent of entrants to teaching, more than double the proportion in any other group, are children of semi-skilled or unskilled workers.

The percentage cited by Anderson and Western for the lower end of the spectrum approximate to that of the Rusden first year cohort in general and to that of the Environmental Studies Department specifically.

Parents education

A greater percentage of environmental studies students' parents (39.1 percent of fathers and 18.6 percent of mothers), had some tertiary education compared with 26.7 percent of fathers and 13.9 percent of mothers for the whole cohort. (See Table 4 and 5 below.) It is of interest to note that in Anderson and Western's (1970, pp.13-14) study, in the aforementioned groups, more parents of students in teaching had less education than those of any other student group in the university.

TABLE 4

Father's education

	Primary/ some secondary	4/5/6 years secondary	some/complete tertiary, non-university	some/complete university
	%	%	%	%
Environmental Studies	43.5	17.4	20.3	18.8
Whole Cohort	45.3	28.0	13.3	13.4

TABLE 5

Mother's education

	Primary/ some secondary	4/5/6 years secondary	some/complete tertiary non-university	some/complete university
	%	%	%	%
Environmental Studies	44.3	37.2	15.7	2.9
Whole Cohort	48.3	37.7	9.1	4.8

Motivations

The responses to the questionnaire 'Motivations for Entering a Teacher Preparation Course' gave an indication as to the probability of students joining the teaching profession for a number of years on completing their course. Twenty-nine percent of environmental studies students indicated on enrolment that they would teach for the whole of their career, compared with 45 percent for the whole cohort and 24 percent felt it was the only job that interested them compared with 32 percent for the whole cohort (see Tables 6 and 7). The results indicate that many students, particularly in the Environmental Studies Department, held a low degree of commitment to teaching or felt ambivalent about a teaching career. The results suggest that some students who enrolled at Rusden SCV did not necessarily consider that they were going to teach in the future. The low commitment to teaching of students who enrol for tertiary education courses is well documented. For example (Blaikie, 1974:91) found that only 29.2 percent of education students at Monash University in 1968 would choose teaching if given a free choice of occupation and at La Trobe University in 1970 Edgar and Edgar (undated) found only 28.4 percent of enrolments would choose teaching if given another chance to choose an occupation. As such a large proportion of education students do not perceive themselves as wholly committed to teaching, it would not seem to augur well for their satisfaction with first year teaching. With the reduced allocation of studentships, the question is raised as to whether there will be an increase or a reduction in the number of students who are motivated to teach. The answer to this type of question may have important implications for the

TABLE 6

Responses recorded by first year students to item 19 by department in which enrolled, SCV Rusden, 1976

Anticipated length of teaching service	DEPARTMENTS																			
	Business Studies		Drama		Environmental Science		Home Economics		Language & Literature		Mathematics		Media Studies		Physical Education		Social Studies		TOTAL	
	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%
– whole of career	22	61	13	41	19	29	54	51	42	63	10	71	9	27	36	45	21	62	226	48
– about 10 years	3	8	-		3	5	5	5	5	8	-		3	9	11	14	1	3	31	7
– about 5 years	–		1	3	4	6	8	8	1	2	-		6	18	3	4	2	6	25	5
– until 3 years bond completed	-		-		1	2	-		2	3	-		1	3	-		-		4	1
– less than 3 years	-		1	3	1	2	1	1	-		-		1	3	-		3	9	7	2
– too uncertain to say	11	31	17	53	38	58	37	35	16	24	4	29	14	41	30	37	7	21	174	37
– no response	–		-		-		-		1	2	-		-		1	1	-		2	
TOTAL	36		32		66		105		67		14		34		81		34		469	

TABLE 7

Responses recorded by first year students to item 20 by departments in which enrolled, SCV Rusden, 1976

TEACHING IS:	DEPARTMENTS																			
	Business Studies		Drama		Environmental Science		Home Economics		Language & Literature		Mathematics		Media Studies		Physical Education		Social Studies		TOTAL	
	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%
The only job which interests me	19	53	6	19	16	24	47	45	19	28	7	50	7	21	23	28	14	41	158	34
One of several jobs I am interested in	16	44	24	75	37	56	52	50	44	66	7	50	26	77	56	69	19	56	281	60
A job in which I am not sure that I am interested in	1	3	2	6	10	15	6	6	4	6	-		1	3	2	3	1	3	27	6
A job in which I am not interested	-		-		2	3	-		-		-		-		-		-		2	
No response	--		-		1	2	--		--		-		-		-		-		1	
TOTAL	36		32		66		105		67		14		34		81		34		469	

development of alternative career paths for students in tertiary institutions which at present emphasize teacher training.

ASAT

The mean score for environmental studies students on the Australian Scholastic Aptitude Test (ASAT) was higher than the mean for the whole cohort (see Table 8). This was also true for the mathematics subscore, whereas the humanities subscore was about the same as for the whole cohort. The results of one way analyses of variance indicated that the differences between the mean scores of departments were significant.

Anderson Score

The mean Anderson Score for environmental studies students was higher than the mean score for the whole group (see Table 8). The results of a one way analysis of variance indicated that the differences between the mean scores of departments were significant. The Anderson Score is the main criterion used for selection of students at Rusden. Since the mean Anderson Scores of departments differ, it follows that a students' choice of course is a major factor in determining whether he is offered a place or not.

Although the Anderson Score was the main criterion used for selection of students, the correlation for the whole cohort between first year performances (based on the aggregate score of 4 subjects) and the Anderson Scores was found to be low (r = .165) and not to be significant. The results therefore called into question the use of the Anderson Score for selection purposes.

HSC Scores

The HSC mean scores of students taking environmental studies were higher (by more than 3 percent) for the mean scores of the whole cohort for HSC biology and geography. However, it is of interest that the mean scores of physical education students in these subjects were higher than the mean scores of environmental studies students. The mean HSC scores analysed by department revealed significant differences in the performance of groups of students at entry. These differences at entry would appear to be partly due to the 'attractiveness' as shown by the preferences for these courses.

There were no significant correlations between HSC biology and geography with their corresponding first year subjects at Rusden. This would suggest both differences in content and methods of assessment between the HSC subjects and the corresponding college subjects.

The HSC subject choice of students

A further point of interest shown in Table 8 is that a number of HSC subjects which are traditionally associated with science, were often not studied by environmental studies students. Of 61 environmental studies students, 56 had studied HSC biology and 42 HSC geography, but only 15 had studied any mathematics, 8 chemistry, 3 physics and 3 geology. So far, our analysis of the *pattern of HSC subjects* attempted by students is incomplete. Nevertheless, it has already been noted that the 8 students who had taken HSC chemistry included those who had taken HSC mathematics, physics and geology. This would seem to endorse the

TABLE 8

ASAT, Anderson and HSC scores of Environmental Studies students and the total first year cohort

	ENVIRONMENTAL STUDIES					TOTAL n = 495				
	N	Mean	SD	Min	Max	N	Mean	SD	Min	Max
ASAT-Humanities	64	28.7	6.3	14	41	461	28.0	5.7	13	43
ASAT-Maths	64	15.4	4.5	3	27	461	13.4	4.9	0	27
ASAT-Total	64	44.1	9.4	17	65	461	41.4	8.5	17	65
Anderson	61	276.1	20.6	220	340	432	272.3	28.4	75	358
HSC-English Expression	60	64.9	10.4	40	85	442	65.3	10.0	35	95
HSC-English Literature	19	62.2	14.5	31	86	202	65.4	11.0	31	91
HSC-Language	4	66.0	10.5	-	-	47	63.3	13.8	33	91
HSC-General Maths	12	66.0	10.1	52	80	111	65.0	13.1	6	98
HSC-Applied Maths	3	49.0	2.0	-	-	29	62.0	13.3	40	86
HSC-Pure Maths	3	53.0	11.5	-	-	40	58.8	15.9	14	87
HSC-Biology	56	66.7	9.2	48	82	251	63.6	11.0	21	91
HSC-Ag. Science	3	68.3	4.7	-	-	5	64.0	7.3	54	72
HSC-Chemistry	8	57.0	9.9	46	71	57	63.5	10.1	44	89
HSC-Physics	3	39.0	6.9	35	47	38	60.9	13.8	34	88
HSC-Geography	42	70.0	10.0	53	92	168	64.6	11.6	32	92
HSC-Geology	3	76.3	7.5	-	-	4	73.3	8.7	64	84
HSC-Australian History	28	64.1	11.8	29	84	196	64.8	11.5	28	96
HSC-Non Australian History	10	62.3	7.6	50	76	94	63.1	11.5	27	89
HSC-Biblical Studies	4	51.5	14.8	-	-	13	57.3	13.8	34	79
HSC-Art	4	62.5	14.5	-	-	63	60.8	11.3	34	81
HSC-Graph. Comm	-	-	-	-	-	6	62.7	14.1	41	81
HSC-Com & Legal	10	65.7	9.1	53	80	98	64.0	11.1	33	95
HSC-Accounting	4	62.3	6.9	-	-	48	62.3	9.7	2	86
HSC-Economics	14	61.64	14.6	35	89	137	65.0	11.5	30	96
HSC-Social Studies	18	60.8	13.4	31	79	129	65.0	11.8	31	88
HSC-Music	1	44.0		-	-	11	56.2	21.1	13	90

view that some students enter the Environmental Studies Department because they do not have the prerequisites for other tertiary courses, whilst others may choose because they are committed to the field of knowledge. Presumably an explanation for the high percentage of first preferences for the environmental studies course was based on students' assumptions that they were not eligible for other courses elsewhere — that is, in the disciplines.

In analyzing the pattern of HSC subjects attempted by environmental science students who enrolled from 1973 to 1975 at Melbourne State College, Elsworth (1976, p.2) reports:

> (i) an increase in the number of students taking Biology, Economics and Geography in later entering years,
> (ii) a concomitant decrease in the number of students taking Chemistry, Physics and Applied Mathematics.

There appears to be an increasing diversification in the subjects which students have studied at HSC and an increase in the perceived importance of biology, economics and geography in the school curriculum.

The diversity of subjects that students have studied at HSC may to some extent account for the low correlation between the Anderson Scores (based on the four best subjects) and the first year performance.

Were the environmental studies students at Rusden those with a superior 'scientific background'? The results in Table 9 below showed that more physical education students had studied mathematics, physics and chemistry and their performances in these subjects was generally superior to those of environmental studies students.

TABLE 9

Comparison of selected HSC subjects' results of Physical Education and Environmental Studies students

Subject	Physical Education (n=85)			Environmental Studies (n=71)		
	N	Mean %	SD	N	Mean %	SD
general mathematics	38	71.4	8.0	12	66.0	10.1
applied mathematics	7	75.4	7.8	3	59.0	2.0
pure mathematics	12	66.4	16.0	3	53.0	11.5
physics	16	67.0	10.0	3	39.0	6.9
chemistry	21	69.1	7.6	8	57.0	9.9

In terms of HSC subjects studied, the physical education students are superior in their performance in most subjects to those who are allegedly specializing in the field.

Since entry students such as those in the Environmental Studies Department cannot all be assumed to have the background knowledge, skills and abilities on which to build, it follows that background material may need to be incorporated into courses so that some students will not be disadvantaged. Alternatively students may require special subject orientation courses or individualized learning materials.

Footnotes

1. Similarly lower figures have been recorded for the Departments of Mathematics (45.5 percent) and Social Studies (51.9 percent).
2. At LaTrobe University 50.1 percent of students have attended State High schools — printout Tertiary Research Centre of the University of New South Wales (Vice Chancellor's Questionnaire) 1976.
3. Interestingly, only 35.4 percent of home economics students come from Melbourne schools.
4. Some fifty percent of postponements were due to economic hardship — information from letters in student records.

References

ANDERSON, D.S. & WESTERN, J.S. Social profiles of students in four professions. *Quarterly Review of Australian Education*, 1970, 3(*4*).

BARNARD, C. The social composition of tertiary students in Australia and the effect of the abolition of fees. *Report No. 4*, Tertiary Education Research Centre, University of New South Wales, 1976.

BARRETT, E.M. A special admissions scheme — mature age unmatriculated students at the University of New South Wales in 1975. Paper given at ANZAAS Conference (Hobart), 1976.

BLAIKIE, N.W.H. Value goals and choice of teaching. In Edgar, D.E. (ed.), *The competent teacher*. Sydney: Angus and Robertson, 1974.

EDGAR, D.E. & EDGAR, P.M. First year teaching: a crisis of morale and competence, occasional paper (undated).

ELSWORTH, G. Report of an evaluative study of the environmental science course, 1975. TERU Report No. 1, 1976.

KINGS, C.B. & DICKINS, R.L. *Special entry and mature students*. Melbourne: Rusden SCV, 1976.

MARSH, B. A report on the academic performance of B.Ed. candidates (La Trobe University) with standard and non-standard qualifications. Paper given at SPATE Conference, Brisbane, 1976.

VAN HELDEN, F.N. Student performance in relation to age. *The Australian University*, 1975, 13(*3*), 262-270.

WALKER, P. The university performance of mature students. *Research in Education*, 1975, 14, 1-13.

WHITE, R.T. From the two cultures to money: trends in enrolment in HSC subjects in Victoria. *Science Education: Research*, 1973, 95-104.

Research in Science Education, 1977, 7, 193-202.

TRANSITION FROM YEAR 12 SCHOOLING TO THE WORKFORCE: DOES SCIENTIFIC ACHIEVEMENT HELP?

R. Lewis

Introduction

There are a number of reasons that may prompt the investigation of an educational problem. Some of these are related to the testing and formulation of theory. Others are more decision oriented in that they aim to provide empirical information that may be relevant to an issue of concern. One issue of current concern is the accountability of educators in schools to the community at large. Recent concern over student competence at the basic skills was evidenced by the public response to investigations of the numeracy and literacy of Australian students, including parliamentary reference* to the findings of the ACER investigation into basic skills in Australian schools. Another area of current community interest is the relevance of school learning to life in the wider social context. Of particular interest is the effectiveness of schooling as a preparation for work experience.

The purpose of schooling is a complex problem. Schools have been assumed responsible for a multitude of functions ranging from the maintenance of social graces, to educating for the ability to structure one's leisure time, or preparation for entry into the workforce (King & Brownell, 1966). Although many educators would agree that the role of the school extends beyond training young people for work, the relative emphasis to be assigned to its other role components would vary with one's philosophical viewpoint.

> The growing concern about the transition from school to work is one aspect of an increasing unease about the social role of schools and other formal educational institutions. (p.4)

The preceding quotation, taken from the Organization for Economic Cooperation and Development (O.E.C.D.) committee's Report on Educational Policy in Australia (1976), draws attention to the relevance of school learning to students' preparation for entry into the workforce. Questioning the need for such relevance does not remove the community concern, nor the usefulness of empirical investigation. The current study does not intend to reflect any philosophical prejudices but rather seeks to provide information on a question of immediate concern, without investigating the validity of the assumptions that prompt the asking of the question.

It is necessary to indicate that the data utilized in this study were not specifically collected with a view to investigating the issue of transition. However, it was possible to carry out some exploratory analyses which help in clarifying relationships between indicators of school performance and indicators of the ease of transition from school to work.

Method

Sample

The data analysed in this investigation were taken from the sample of secondary students who participated, during 1970, in a survey of achievement in Science in Australia and who were, at that time, in their final year of secondary schooling. Australia was one of 22 countries which participated in the series of surveys conducted by the International Association for the Evaluation of Educational Achievement (IEA). The Australian Council for Educational Research (ACER) was the National Centre responsible for compiling Australian data. The sample under survey in this study represents a sub-set of the total Australian sample of 4201 students from 194 schools about whom data are held on file in the IEA data bank at ACER. The schools within each state which were sampled in this study were selected with a probability proportional to size and the sampling stratified by school type and location. Students were selected at random from within each school and the sample of students can be considered to be representative of students in 1970 in their final year of secondary schooling in New South Wales, Victoria, Queensland, South Australia, Western Australia and Tasmania. The further interrogation of this sample in 1972 was undertaken by M.J. Rosier and K.N. Ross as part of a follow-up study. The information collected in the follow-up study has provided the criterion measures examined in this report.

---

*House of Representatives Daily Hansard. A.G.P.S. Thursday, 14 October 1976, p.1846.

The section of the follow-up questionnaire dealing with employment requested information on the respondent's main occupation and the number of months during which the occupation was undertaken. The respondent also indicated if employment was on a full-time or part-time basis. The sample selected for study in this report contained all respondents who indicated some full-time work. This definition included both those who continued with part-time study and those who ceased to study. However the findings will be presented separately for respondents who did not engage in further education and those who did. This will allow an investigation of the relationship between continuation of education and the independent variables.

Variables

Independent variables

The independent variables were restricted to only those achievement measures which were scaled identically in each State. This restriction was caused by a desire to use the total Australian sample thereby ensuring sufficient numbers of respondents to allow meaningful generalizations. The tests used were the science subtests incorporated into the IEA Science test. These were measures of performance at Physics, Chemistry, Biology and ability in practical work. With the exception of the last named test (which had 12 items) the achievement measures all were 16 items in length. The correlations between the IEA achievement tests and Year 12 achievement tests in the sciences ranged between .35 and .62 with a median value of .55 and were all highly significant. Therefore the IEA science tests appear to be very good indicators of school achievement in the sciences.

Dependent variables

Three work-related measures were used as criteria for investigation. These were: (1) the type of employment undertaken after completing the final year of secondary school; (2) the difficulty encountered in obtaining employment, and (3) the retention of the same position after one year's employment.

Results

Type of employment

This variable was initially coded into 99 categories using the Broom, Jones and Zubrycki scale (1968) as a basis for definition of type of employment. Its application, however, resulted in the identification of nine major types of occupation undertaken by students who had completed their final year of secondary schooling. These are described in Table 1, which indicates the proportion of Australian students entering different occupations. The data are presented separately for respondents studying part-time and those who ceased to study completely.

TABLE 1

Type of employment classified by time at study

Percentages recorded Occupation	Australian Respondents*	
	Part-time Students N = 454	Non Students N = 421
White Collar		
Nurses	24	1
Technicians	22	9
Cashiers	9	5
Clerks	20	49
Public Servants	6	9
Other White Collar	5	8
Blue Collar		
Shop Assistants	2	4
Labourers	0	1
Other Blue Collar	13	15
Total	101%	101%

*The data are weighted by state.

Workers who studied part-time were significantly more likely ($p < 0.001$) to be working as technicians or nurses than those who didn't study, nurses being by definition part-time students. Almost half of the workers who were studying part-time were nurses or technicians. Another 20 per cent were clerks and only 15 per cent were classified as blue collar workers.

The occupation most frequently chosen by non student workers was clerical work. This category contained almost half of the non studying workers. It is interesting to note that the proportion of blue collar workers was not significantly related ($p > 0.05$) to part-time study, the difference between studying and non studying groups being 5 per cent in favour of the latter group.

Before concluding considerations of the type of employment undertaken by students who have completed Year 12 it is necessary to indicate the existence of a large group of non respondents (N = 1305). A comparison of mean scores on the independent variables indicated that this group of non respondents generally had lower attitudinal and achievement scores than the group of full-time workers and were best described as lower achieving students. It is probable that the non respondents would be more likely to be working in blue collar occupations than either of the groups of workers for which data are presented in Table 1. If, for example, 50 per cent of the non respondents engaged in blue collar occupations, the best estimate of blue collar workers for all students leaving secondary school after completing their final year (excluding full-time bonded students) increases from approximately 17 to 40 per cent.

Difficulty in obtaining employment

The majority of Year 12 Australian students progressing to full-time employment reported no difficulty in obtaining a job. Those who did find it difficult indicated a variety of reasons listed below:

(i) lack of suitable experience/qualifications;
(ii) job didn't suit (hours, location);
(iii) over qualified (age, educational level);
(iv) inappropriate outlook (lack of motivation, determination);
(v) bias by employer (students, females, foreigners);
(vi) no job available;
(vii) unclassified.

However, since only about 30 per cent of the respondents not studying concurrently and 20 per cent of the full-time workers studying part-time had any difficulty in obtaining jobs, the fragmentation of their responses into seven categories was avoided. The difficulties were dichotomised. Students were considered as experiencing difficulties of a personal nature [categories (i) to (iv) above] or difficulties beyond the control of the applicant [categories (v) to (vii) above]. The proportions of students in these categories are recorded in Table 2, both for students who studied part-time and those who did not study.

The non students had significantly greater difficulty ($p < 0.001$) in finding employment than did part-time students, owing to increased proportions indicating both types of difficulties in the former group of students.

The data in Table 2 relating Difficulty and Time at Study could indicate that the type of work which fosters further part-time study was easier for Year 12 students to obtain. It may alternatively be interpreted as indicating that superior respondents would tend to continue with part-time study and that these superior respondents would also be least likely to have difficulty impressing prospective employers. However, data presented later in this paper, relating cognitive performance to difficulty in obtaining employment indicates that the latter alternative is relatively unsound. Therefore it may be hypothesised that employers seeking candidates for jobs unrelated to further education tended to discriminate against applicants who had completed their secondary schooling more so than those who offered work which encouraged further training (e.g. trade school).

Stability of employment

This variable was used to indicate if the respondents retained their employment for

TABLE 2

Difficulty and stability of employment classified by time of study

Percentages recorded	Australian Respondents	
	Part-time Students	Non Students
Difficulty		
Employee defined difficulty	13	18
Employer defined difficulty	5	12
No difficulty	83	70
Stability		
Changed job	11	27
Returned to study	3	10
No change	86	64

their initial year in the workforce. It acted as an indicator of suitability of employment. Implicit in the use of this indicator is an assumption that the student who has most successfully made the transition from secondary education into the workforce is one who has no desire or need to change jobs. The three coding categories used for Stability were (1) no change in employment; (2) changed jobs, and (3) returned to full-time study. Table 2 indicates the proportions of students in these response categories.

The stability of employment of respondents who continued with part-time study was significantly greater ($p < 0.01$) than for the non-studying workers. This was accompanied by a significantly greater proportion ($p < 0.01$) of non-studying workers who changed jobs in their first year of employment. One possible explanation for this finding could be that jobs with opportunity for part-time study provided greater satisfaction and less likelihood of change. A second possibility could be based on an assumption that superior students were the ones who maintained part-time study (irrespective of the opportunities provided by their occupation) and these superior students were most likely to obtain more satisfying jobs. A third alternative explanation could be based on the assumption that many employers who provided opportunity for part-time study also bonded employees, making job stability more likely.

The next step in the analysis involved relating performance in the sciences to the type of jobs obtained. For each of the science tests, respondents were divided into five or more approximately equal groups on the basis of their test scores. Then the proportion of respondents in each group who obtained work as nurses was calculated. This procedure was repeated for each of the occupations. Tables 3 and 4 record all data which showed any evidence of linear trends between any achievement test and any occupation.

For students studying part-time, greater performance on all Science subjects was associated with a greater likelihood of technical occupations. With the exception of Biology and Chemistry the same relationship also applied between achievement and the 'other blue collar' occupations (skilled trades). The opposite relationship held between achievement in Chemistry, Physics or Practical and Nursing, with greater achievement corresponding to less likelihood of being a nurse. Clerical work was the only other occupation which related to

TABLE 3

Type of employment classified by achievament in the sciences for part-time students

Biology Score	0-4	5-6	7-8	9	10	11-12	13-14
Technicians (%)	15	16	9	23	25	42	25
Chemistry Score	0-2	3-4	5-6	7-8	9-10	11-12	13-14
Nurses (%)	38	31	19	16	15	5	10
Technicians (%)	3	13	30	23	38	45	52
Physics Score	0-2	3-4	5-6	7-8	9-10	11-12	13-4
Nurses (%)	40	44	24	15	12	5	14
Technicians (%)	2	7	21	25	31	30	73
Other Blue Collar (%)	10	6	12	8	23	20	5
Practical Test Score	1	2	3-4	5	6	7-8	9-10
Nurses (%)	22	35	37	20	14	18	10
Technicians (%)	0	8	13	20	28	31	50
Clerks (%)	39	24	17	22	20	18	8
Other Blue Collar (%)	6	11	10	13	17	18	15

TABLE 4

Type of employment classified by achievement in the sciences for non students

Biology Score	0-4	5-6	7-8	9	10	11-12	13-14
Clerks (%)	52	53	54	51	36	56	41
Other Blue Collar (%)	12	9	16	12	22	16	24
Chemistry Score	0-2	3-4	5-6	7-8	9-10	11-12	
Other Blue Collar (%)	10	15	19	11	23	28	
Physics Score	0-2	3-4	5-6	7-8	9-10	11-12	
Clerks (%)	63	52	54	39	49	22	
Other Blue Collar (%)	5	14	9	24	18	39	
Practical Score	1	2	3-4	5	6	7-8	9-10
Clerks (%)	74	49	49	56	48	43	36
Other Blue Collar (%)	0	11	12	19	20	16	28

achievement for part-time students and this related negatively to respondents' scores on the practical test. Less relationships were noted for the group of non students (Table 4). Greater performance on each of the Science tests indicated a greater likelihood of one of the 'other blue collar' jobs and (with the exception of Chemistry) lesser likelihood of clerical work.

The major difficulty in attempting to interpret these findings is the confounding effect of student sex. The data in Table 5 indicate that most of the relationships described in Tables 3 and 4 can be adequately explained in terms of student sex, if one assumes that males outperform females on the IEA science tests. Such an assumption would be supported by the predominance of males studying sciences at Year 12 level.

The difficulty in obtaining employment was related to achievement in the sciences by considering achievement at three levels and recording the proportions indicating difficulties within each achievement level. The data are recorded in Table 6.

TABLE 5

Type of work classified by student sex

	Part-time Students		Non Students	
	N = 278 Male	N = 171 Female	N = 247 Male	N = 172 Female
Nurses	1	60	0	2
Technicians	30	10	6	11
Cashiers	12	4	5	3
Clerks	22	16	43	58
Public Servants	8	3	9	8
Other White Collar	6	4	11	5
Shop Assistants	1	2	3	4
Labourers	0	0	2	0
Other Blue Collar	20	1	21	9
Total	100%	100%	100%	100%

The only statistically significant difference between the proportion of high and low achievers reporting no difficulty occurs for the non students' practical scores ($p = 0.075$). Therefore, although the type of job obtained was related to achievement in the sciences, the difficulty experienced in obtaining work was not. Further analyses (Lewis, 1976) indicated that achievement in the sciences was also generally unrelated to the difficulty encountered in obtaining employment as Technicians, Clerks, Nurses, or 'Other blue collar workers'. Therefore the varying difficulties involved in obtaining different categories of jobs may not be considered responsible for the finding.

Three interpretations of these findings seem plausible. The first would have as its basic assumption, the independence of the science skills and those skills sought by prospective employers. The second would propose a severe lack of guidance in schools as to the type of jobs available, the conditions of employment and the qualifications required. The third would

TABLE 6

Difficulty in obtaining employment classified by achievement in the sciences

Percentages recorded Difficulty	Part-time Students				Non Students			
	Employee Defined	Employer Defined	None	N	Employee Defined	Employer Defined	None	N
Biology								
Low	11	3	86	102	25	12	65	124
Middle	17	3	80	227	17	8	75	136
High	10	3	87	110	23	14	63	142
Chemistry								
Low	11	3	86	95	16	12	72	101
Middle	14	6	80	247	20	11	69	217
High	11	2	87	97	16	13	71	84
Physics								
Low	12	3	85	132	18	9	72	149
Middle	14	6	80	148	13	16	71	92
High	12	4	84	159	22	11	67	161
Practical								
Low	16	6	77	101	12	11	76	132
Middle	10	4	87	236	20	10	69	126
High	17	6	77	102	21	15	63	144

be related to the second and suggest that the level of complexity of employment sought by respondents varies in accord with their ability and achievement.

Although the first explanation is probably valid to some extent, it is the latter two explanations that are the most likely causes of the lack of relationship between science achievement and the difficulty of entry into the workforce. The recent moves in Victoria to improve the quality of career counselling in schools and to train selected teachers for this task seems to indicate an assumption that career counselling can assist in reducing the level of difficulty students encountered by guiding them to employment of a 'suitable' type and level of complexity.

However, one may accept the findings as an outcome of the third explanation without recognising a need for guidance and counselling. It is plausible to accept that a 'natural law' exists which results in a proportion of students with high expectations attempting to 'aim high' in terms of employment. Therefore it would not be the result of a lack of information that led to the student encountering difficulty and it may not, and perhaps should not be within the role of the school to 'remedy' the situation by lowering the student's sights. A study of the nature of this current investigation cannot distinguish between intentional and purposeful mismatch of job candidates' ability and job complexity. It can and does, however, indicate a strong need for further investigation of the relationship between science achievement, career counselling, work expectations and employment sought.

Table 7 presents data on the stability of employment categorised by level of achievement in the sciences.

TABLE 7

Stability of employment classified by achievement in the sciences

Percentages Recorded	Part-time Students				Non Students			
	N	Changed Job	Work to Study	No Change	N	Changed Job	Work to Study	No Change
Biology								
Low	96	11	4	84	113	25	11	65
Middle	218	13	3	84	133	26	8	67
High	103	7	2	91	139	29	9	62
Chemistry								
Low	86	9	2	88	111	25	8	67
Middle	236	13	2	85	177	31	9	60
High	95	8	5	86	97	21	10	69
Physics								
Low	125	12	2	87	146	25	13	62
Middle	137	18	4	78	90	28	2	70
High	155	6	3	91	149	28	9	61
Practical								
Low	96	15	4	81	103	28	13	59
Middle	227	11	3	86	145	27	5	68
High	94	9	2	89	137	26	11	64

There are no significant differences between the stability of high and low performers. Once again the design of the current study can do little to clarify the causes of such findings. All that can be said by way of description is that science achievement is unrelated to the likelihood of the respondent retaining his or her initial job for one year.

The aim of this investigation was to provide empirically-based information on the relationship between scientific achievement and transition from Year 12 schooling to the workforce. The findings indicate that science achievement was related to type of work undertaken but independent of the difficulty encountered in obtaining employment and the stability of such employment. Independence does not necessarily imply irrelevance but may be interpreted as suggestive of the need for career counselling. Irrespective of interpretation, the findings do appear to warrant further investigation of the relationship between science performance and transition, if only to prevent simplistic interpretations of a complex problem.

References

KING, A.R. JR. & BROWNELL, J.A. *The curriculum and the disciplines of knowledge.* N.Y.: John Wiley and Son, 1966.

LEWIS, R. *Transition from School to Post Secondary Education in Australia IEA (Australia) Report 1976:1.* Hawthorn, Victoria: ACER, 1976.

O.E.C.D. *Report on educational policy in Australia.* Canberra: A.G.P.S., 1976.

Research in Science Education, 1977, 7, 203-211.

STUDENT TEACHER QUESTIONING IN TWO ENQUIRY DIRECTED SETTINGS OF SCIENCE TEACHING

R. A. Fawns

This investigation reflects an interest in the manner in which relatively inexperienced teachers perceive and use laboratory-based curricula packages, such as the A.S.E.P. units. In particular, I am interested in building up a picture of the way in which young teachers attempt, through questions asked, to establish and sustain the hypothetical or enquiry mode in two teaching settings characteristic of science classrooms, and the way in which the two teaching settings may influence these attempts. These settings are:

(a) the small group activity and an ensuing class discussion;
(b) the demonstration-discussion context.

In focussing upon styles of verbal interaction between young teachers and their pupils, and the activity based materials, the study will also be concerned with how language is perceived and used as an instrument of teaching. The data reported here are part of a larger project (Fawns, 1976).

Design

The exploratory study examines the questioning of 6 student teachers in a single "complex" science lesson which required the student teacher to arrange a number of learning contexts in the one period for a grade 8 class. The lesson plan was based around two pages, 78 and 79 from the ASEP unit, *Forces*.

The design was broadly of the post-test only control group form with randomization handled in large part by the distribution of mixed ordered teaching plans.

An attempt was made to devise a design and procedure which permitted experimental manipulation of classroom variables, in this case the sequencing of classroom settings, so that a degree of control was exercised over lesson material and the setting in which it was taught, but without the degree of detailed scripting employed in the Christchurch studies (Church, 1971). Decisions about the appropriate level of laboratory type control or scripting over the independent variables could ideally depend as much upon the nature of the particular learning-teaching settings being investigated as upon methodological stance. It seemed appropriate that if the scenarios were standard forms well understood by teacher and pupils, detailed scripting in such studies would be less necessary. (The dependent variable was student teacher questioning behaviour and the setting was manipulated as the independent variable.)

The lesson content and conceptual sequence was the same in all cases, but the teaching strategy sequence was altered from the published format for a randomly allocated half of the lessons (see Fig. 1).

In the data reported here, student teachers of one training programme were used. As well as a direct coding procedure based on Bloom, a transcript was made of each lesson to which the Smith and Meux (*"The Logic of the Classroom"*), and the Withall (*"Socio-Emotional Climate Index"*) systems were applied.

	Phase I	Phase II	Phase III	Phase IV
Content	The phenomenon of floating	Displacement and upthrust correlated	Upthrust in fresh and salt water	Practical application of Plimsoll lines
Teaching Sequence I: (published sequence)	Introduction	group activity-discussion	activity demonstration-discussion	discussion of second hand data
Teaching Sequence II: (modified sequence)	Introduction	demonstration-discussion	group activity-discussion	discussion of second hand data

Only Phases II and III were manipulated.

FIGURE 1

Lesson summary — content and teaching sequences

The instruments

A. The Smith and Meux Instrument "The Logic of the Classroom"

The Abstract Constructs of "High and Low Enquiry": In this study, the division of the Smith and Meux categories into those making "low enquiry" and those making "high enquiry" demands on pupils was made on the assumption that questions in certain categories (describing, designating, stating, reporting and defining) will usually have required only recall of "scientific facts", whilst the questions classified in the other categories (substituting, evaluating, opining, classifying, comparing conditional inferring and explaining) will usually require processes of thought which are more logically complex and hypothetical.

In the Ladd and Anderson study (1970), substantial agreement was obtained between sixteen science educators when they were asked to classify the twelve functional question categories of the Smith and Meux instrument, according to those which would require high enquiry pupil responses and those which would require low enquiry pupil responses. The clusters of categories which emerged from their study as operational descriptions of both high and low enquiry seem to correspond closely to the clusters inferred from the ratings of the student teachers who participated in this study. The Ladd and Anderson groupings were adopted in this study. Categories 1-5 describe low enquiry and categories 6-12 describe high enquiry processes. Some limited comparisons were attempted with the Tisher (1970) Brisbane data.

B. The Social-Emotional Climate and Student Teacher Questioning

The system employed here is the Withall (1951) *Social-Emotional Climate Index* which has been used in research, teacher-training and supervision. It is restricted to verbal interaction. The coding unit in Withall's study was described as "word or words, sentence or sentences, containing an idea or theme and a dominant intent aimed at either teacher or pupil support".

Withall assumed that his seven categories lay along a continuum of high-low learner supportiveness or low-high teacher-centredness and that if a majority of teacher statements were classified in the first four categories, the climate would be said to be "learner-centred", and if the majority of statements were classified in the final three categories, the climate was said to be "teacher-centred". Comparisons with the Brisbane study were again attempted.

Data processing

A transcript of each lesson was prepared and the tape replayed in conjunction with the coding of the questions by three independent coders in the main study. The coders were trained graduate teachers.

Statistical analysis

(a) Reliabilities

In this study the reliabilities quoted were percentage agreement on each item coded — for the cognitive dimension .75 and for the social-emotional climate .70. For the category totals the reliabilities were in the range .75 to .87.

(b) Hypotheses Testing

The hypotheses were tested using a two-way analysis of variance with repeated measures (Winer, 1971).

Results and discussion

A1. The logical and enquiry intent of student teachers' questions

The group distribution (total) can be compared with that reported by Nuthall and Lawrence (1966), and also Tisher (1970), where lessons were also to pupils of grade eight, but in which trained teachers were observed conducting descriptive lessons in which little primary data was generated or discussed.

The high frequency of management questions in the Melbourne study could be a student teacher effect arising from their relative inexperience in teaching, particularly of complex lessons involving the management of a number of different learning contexts. It is also possible that the high number of such questions reflected their low institutional status in the school. However, few of the management questions asked by this group were basically disciplinary in intent; most were directed to the task. Moreover, these student teachers were in the final weeks of an intensive programme of school practice. They had been teaching in the school in which they taught the lesson for at least a term, and had been involved in continuous practice on at least two days per week for the academic year. Whilst they may be regarded on many criteria as young teachers, they were still student teachers in the sense that these classes were not "their own" and this could have systematically influenced the patterns

TABLE 1

Comparison with other studies of distribution of logical operations and levels of enquiry intended by teacher question*

Smith and Meux categories	Nuthall and Lawrence (1966) Classes in arithmetic, languages and social studies	Tisher's Brisbane study Classes in science		Current Melbourne study Student teacher's Classes in science	
1. Describing	26	29.7	75.0	3.2	26.2
2. Designating	11	32.5		8.1	
3. Stating	12	10.7		3.1	
4. Reporting	2	0.0		10.1	
5. Defining	5	2.1		1.7	
6. Substituting	2	0.3	20.4	0.1	22.2
7. Evaluating	4	0.4		1.2	
8. Opining	2	0.7		3.6	
9. Classifying	2	0.8		0.1	
10. Comparing & Contrasting	2	2.4		2.1	
11. Conditional inferring	7	7.2		4.2	
12. Explaining	11	8.6		10.9	
13. Management (a) and (b)	3	4.6		51.1	
Total Questions	"low"	1800		823	

*Ladd & Anderson Categories 1-5 "low enquiry"/Categories 6-12 "high enquiry".

of enquiry and the social climate to some degree. It seems likely that at least part of the difference between the Brisbane and the Melbourne studies is attributable to the difference in teacher-programme intent – the descriptive versus the enquiry-based curricula.

In these two studies there seems to be little difference in the proportion of higher level (categories 6-12) to total questions and, while a smaller proportion of low level questions may have been asked in the interpretation of primary data, there was a proportionate increase in the number of management and directive questions used in this environment. Pupils' responses to questions coded in the category "Explanation" were examined using the Nuthall and Lawrence (1966) subscriptions. Irrelevant and descriptive responses constituted 64% of the responses and a fairly high proportion of these were accepted in the interests of continuity.

The implication is that the nature of scientific explanation was poorly modelled in these classes and that often the pupils either did not recognize or understand what was expected.

A2. Analysis of 'enquiry' orientation of questions asked in two class room settings

One major aim of the experimental studies was to investigate the influence of the two teaching settings upon the promotion of high level enquiry in the teacher questions. The results of the analysis of the proportional data is presented in Tables 2 and 3 in which student teachers are grouped on teaching sequences.

TABLE 2

Proportion of all questions asked by student teachers in which the proportion of high level enquiry was intended

	Teaching Setting		
	Demonstration - discussion	Activity group discussion	Mean
Teaching sequence I (N = 3) (Group activity-discussion followed by demonstration-discussion)	0.493	0.183	0.339
Teaching sequence II (N = 3) (Demonstration-discussion followed by group activity-discussion)	0.350	0.093	0.222
Mean	0.422	0.139	

SOURCE	*M.S.*	*dF*	*F*	*P*
Total	348.909	11		
Between students	146.200	5		
Between sequences	408.333	1	5.062	0.087
Error 1	80.667	4		
Within students	517.833	6		
Between settings	2408.333	1	14.222	0.021*
Sequence x setting	21.333	1	.126	0.7371
Error 2	169.333	4		

TABLE 3

Proportion of all questions asked by student teachers in which management and continuity of class activity was the predominant intent

(Category 13a – public questions in which there was no logical demand)

	Teaching Setting		
	Demonstration-discussion	Activity group discussion	Mean
Teaching sequence I (N = 3)	0.153	0.407	0.280
Teaching sequence II (N = 3)	0.287	0.636	0.462
Mean	0.220	0.522	

ANOVA on proportion of all questions in which the teacher intent was management and continuity

SOURCE	*M.S.*	*dF*	*F*	*P*
Total	490.629	11		
Between students	226.283	5		
Between sequences	990.083	1	28.021	0.008*
Error 1	35.333	4		
Within students	710.917	6		
Between settings	2730.083	1	7.452	0.053
Sequence x setting	20.083	1	0.091	0.664
Error 2	366.333	4		

The Table 2 analysis indicates that, in the context of the demonstration and discussion, a significantly higher proportion of questions intended to promote high level of enquiry was asked than in the group activity and discussion setting (F = 14.22, $p < .05$). No significant differences were found between sequences or settings in the proportion of questions in which low level enquiry was being promoted.

The Table 3 analysis suggests that the published A.S.E.P. teaching required less overt management from the teacher than Sequence II in which the worksheet I was taught by demonstration and worksheet II by class activity.

Although marginally outside the .05 level of significance, there was a difference between the means for teaching settings consistent with that found in a preliminary study (Fawns, 1976), more "management" questions being asked in the class activity setting than in the demonstration setting.

An analysis of the combined 13a (public) and 13b (private) categories of management and continuity indicated that there was a significantly higher proportion of management type questions asked in the group activity class discussion setting than in the demonstration-discussion setting.

B1. Analysis by categories of learner and teacher centredness

The same student teacher questions were categorized for the social-emotional climate they were intended to support. Table 4 shows the distribution of questions for each student teacher and compares the combined frequencies with those of Tisher (1970).

TABLE 4

The distribution of questions asked by student teachers into categories of the Withall social-emotional climate index

	Withall	Student						Melbourne study		Brisbane study	
	Category	1	2	3	4	5	6	mean %		mean %	
Agreeing/ supporting	1	0.7	0.0	1.1	0.0	0.0	0.9	0.1	51	6	37
Clarifying	2	13.1	4.0	14.4	10.4	7.3	8.3	9.6		8	
Problem structuring	3	38.6	37.1	34.4	56.2	29.1	38.2	41.3		23	
Administrating	4	17.9	29.8	16.6	14.6	32.7	34.2	24.2	49	47	63
Directing	5	24.8	21.0	21.1	14.6	10.9	16.6	18.2		10	
Reproving	6	3.4	8.0	12.2	4.1	20.0	1.8	5.9		6	
	Total	100	100	100	100	100	100	100		100	
	No. of questions	145	124	90	192	55	217	823		1800	

Whereas by the Withall criterion all the Brisbane classrooms were classified as "teacher centred" (with a range of 54 to 71 percent of statements by the teacher classified in categories 4-6), in the Melbourne lessons three out of six were classified as "teacher centred", with 33 to 64 percent of questions classified in the same categories. However, the meaning of these labels and the basic assumptions of the instrument are somewhat confused as the Melbourne classes appear more "teacher centred" and less "learner supportive" when the total proportions in categories 5 and 6, and 1 and 2 are compared with the Brisbane sample. As a result of concern

for validity, only the extreme Withall categories, 1 and 2, and 5 and 6, were used in the experimental studies. Lower frequencies and proportions were analysed and no significant differences between the teaching settings or sequences were found. However, significant sequence setting effects were found for both "positive learner supportive" (1 and 2) and "teacher centred" (5 and 6) questions.
($F = 13.5$, $p < 0.05$ and 17.8, $< p$ 0.05, respectively)

These analyses suggest that significantly more "teacher centred" and significantly fewer "positive learner supportive" questions were asked in dealing with the second worksheet, towards the end of the lesson, than with the first work sheet. This may result from the position in the lesson or from the nature of material in the work sheets.

Conclusions and inferences

Whilst this exploratory study of the language of young teachers cannot be described as an evaluation of the ASEP *Forces* unit, the results support the hypothesis that in the teaching of enquiry-based programmes such as ASEP, at least with junior classes, young teachers are less likely to perceive or present science as a "rhetoric of conclusions". The processes of memory and recall of factual knowledge appear to have been effectively de-emphasized in both the materials and classroom interaction. However, the same young teachers appear to find difficulty in sustaining the "narrative of enquiry" emphasized in ASEP programmes. Problems in the logical development of the lessons studied seem to arise both in guiding pupil observation and relating these often incomplete and disparate observations to the development of the organizing ideas or concepts in the materials.

In classroom discussion, the nature of scientific explanation was often poorly modelled pupils' responses indicated that they did not recognize or understand what was expected when they were asked for a scientific explanation. It appears that the proportion of unacceptable responses to such complex logical questions could be higher in junior science classes of young teachers where the active curricula is in use.

Within the framework of developmental-interactionist theory of instruction, the group activity is presumed to permit direct enactive experience with the objects of knowledge, whilst the discussion during and after is to facilitate the development of the iconic and symbolic meanings of the experiments through descriptions, analyses and/or verification of the objects of knowledge. Appropriate discussion in these terms was often attempted, but frequently concluded with the teacher, directly or indirectly, telling her pupils what should have been observed and inferred.

The logical management problems can also be inferred from the prevalence of simple incidents in the typed transcripts. Complex incidents of the type described by Nuthall and Lawrence (1966) involving the answering of subsidiary questions within the context of one major question were far less frequent.

Whether the situation described here can be said to be the general pattern of young science teachers using ASEP programmes at junior level is an open question.

The results support the view that the teaching settings may be important in shaping verbal interactions between teacher and pupil and suggests that Gump's emphasis on the "behavioural setting" as an appropriate unit of observation and explanation deserves further investigation (Gump, 1971). He suggests that a comprehensive picture of classroom phenomena necessary to establish a theory of the teaching process can be given in terms of the organization and qualities of behavioural settings. Such a theory may well be appropriate to a subject, like junior science, taught in few widely used and rather standard and formal settings. The repeated-measures design may offer interesting possibilities for future research into these settings. The findings of this study have implications for research into teacher questioning. They illustrate the importance not only of controlling the lesson material in order to attribute variance in question asking to the teacher, but also the need to consider the influence of the teaching contexts within the lesson upon the nature, distribution and patterning of the questions. The data imply that attention should be given in teacher training to the questions students ask, particularly in laboratory and post-laboratory discussions in the management of concept development in activity-based programmes. There seemed to be a considerable optimistic reliance by the young teachers on the published material to create not only interest but also to structure through direct experience the cognitive conflict necessary to facilitate conceptual reorganization and development.

For the newly-trained graduate teacher, first year out, and in the isolated and often embattled science classroom, an early effect of inadequate preparation for this praxis may well be a personal crisis of confidence in the educational value of enquiry-based modes of instruction, followed by a reversion to the more familiar didactic mode, and an eventual rejection of the underlying progressive rationale.

Mini-courses have been widely advocated for the improvement of specific questioning skills. However, one might question the ideological, logical or empirical basis of the teacher training formulated on the particular imperatives of these instruments alone. In this sense, the following critical questions could be asked of skill-based programmes of teacher education.

(a) To what extent can a teaching skill be identified as critical in the sense that it is regarded as the only effective means of achieving a particular learning outcome?

(b) To what extent does the programme place restraints upon the teacher's freedom to select settings and purposes?

(c) To what extent does the programme removed from the teacher the responsibility for deciding what sort of teacher she should be in the classroom?

The second and third questions imply that a high value should be placed on preparing the teacher for autonomy in "the age of decision and choice" in science teaching. Perhaps the most significant contribution we can make is to encourage the development of a personal critique of each student's practice which makes their own teaching the object of continuous enquiry. (Newsome and Fawns, 1975). This study implies that a valuable part of such a contribution may be a clinical analysis of the nature, potential and limitations of the settings in which junior science is taught.

Somewhat different processes are involved in enquiry as it takes place in the various settings manipulated by science teachers and this study suggests that we should perhaps be choosing to support the emphasis we place on activity in different terms than has been examined here. However, generally there is the need to research the effects of the emphasis placed on "activity" in our new science materials (Yager, 1969). In this study, few ideas, speculations or personal theories were offered by the pupils in any setting. In fact, few offerings were longer than five words. It seems likely that the popular belief that 'current activity is future thinking' is a limited view and the claim made by Barnes (1969) amongst others, 'that present talking is future thinking', deserves closer examination. If language is used as an instrument of classroom management rather than as an instrument of learning, the development of thought may well be inhibited by inadequate and inappropriate language practice.

References

BARNES, D. *Language, the learner and the school.* Harmondsworth: Penguin, 1969.

CHURCH, R.J. The effects of systematic changes in standard for science lessons. *Educational Research Newsletter*, 1971, 4, 15-30.

FAWNS, R.A. Student teacher questioning in two enquiry directed settings of science teaching. Unpublished M.Ed. thesis, University of Melbourne, 1976.

GUMP, P. What's happening in the elementary classroom, in Westbury, I. and Bellack, A. (eds.), *Research into classroom processes*, New York: Teachers' College Press, Columbia University, 1971.

LADD, G. & ANDERSON, H. Determining the level of enquiry in teachers' questions. *Journal of Research in Science Teaching*, 1970, 7, 395-400.

NEWSOME, B. & FAWNS, R.A. Theory and practice in teacher education – a developmental-interactionist view. In Orton, T. and Macmillan, Clec, *Course B: An alternative in teacher preparation at the University of Melbourne. An evaluative case-study.* Melbourne: Centre for the Study of Higher Education, University of Melbourne, Parkville, 1976.

NUTHALL, G. & LAWRENCE, P. *Thinking in the classroom.* Wellington: New Zealand Council for Educational Research, 1966.

SMITH, B.O. & MEUX, M. *A study of the logic of teaching*, U.S. Department of Health Education and Welfare, Office of Education, Orbaner, Bureau of Educational Research: University of Illinois, 1962.

TISHER, R.P. *A study of verbal interaction in science classes and its association with pupils' understanding in science,* University of Queensland Papers, Faculty of Education, 1970, 1-9.

WINER, B.J. *Statistical principles in experimental design*, New York: McGraw Hill, 1972.

YAGER, R.E., ENGEN, H.B. & SNIDER, B.C.F. Effects of the laboratory and demonstration methods upon the outcomes of instruction in secondary biology. *Journal of Research in Science Teaching*, 1969, 6, 76-86.

Research in Science Education, 1977, 7, 213-222.

THE IMPACT OF SCIENCE FACILITIES ON SCIENCE LEARNING

John Ainley

The study

Under the Science Facilities Program (sometimes known as the Commonwealth Science Scheme) some $124 million was provided for the improvement of science facilities in secondary schools. This money was spent on both laboratories and on apparatus and was applied in government and non-government schools.

While the prime purpose of this Program was to improve physical facilities for teaching science it is clear that it was intended that improved facilities would result in 'better' teaching or an improvement in the quality of science education (Menzies, CPD, 1964, 1939; Fairbairn, CPD, 1971, 2665). This paper reports an examination of the impact of science facilities on teaching and learning practices in science. It forms part of a larger study of the Australian Science Facilities Program.

Information about what happened in science lessons was obtained by means of a questionnaire completed by students. The questionnaire was based on the objectives of the Commonwealth Government's Science Facilities Committee. As the objectives involved both the environment in which science was studied and the activities employed, these two aspects formed the basis of the questionnaire.

The Environment for Learning about Science

In this section of the questionnaire three concepts concerning the learning environment were stressed: involvement, organization, and stimulation through variety. These corresponded with three principal concerns of the Science Facilities Committee.

1. The views of the committee about the design and provision of science rooms were based on a belief that *individual involvement* in science learning was valuable. It particularly related this view to the need to accommodate new curricula, to allow students to do experiments themselves, and to foster participation in group discussions (Commonwealth of Australia, 1976, 15; 1973, 11).
2. A second premise was that the learning environment should be *well organized* so that students could work effectively. Published design brochures gave considerable attention to space requirements of rooms in which orderly work and effective management of resources were possible.
3. The third concern of the Committee was that science lessons should be *stimulating through the variety* of methods used. Stimulation was seen as coming through two facets of variety: through the use of varied and flexible instruction and in the use of a wide variety of stimulating materials in the room. In its final report the Committee stated that a science room should be suitable for all kinds of activity including, *inter alia*, demonstrations, pupil experiments, group discussions, films, or a mixture of activities at the one time (Commonwealth of Australia, 1976, 16).

These aspects of the learning environment had been used in other studies of teaching behaviour and classroom process variables. For example one study of some 6000 teachers (Ryans, 1960) identified three patterns of teacher behaviour.

Pattern X — warm, understanding, friendly *versus* aloof, egocentric, restricted behaviour

Pattern Y — responsible, businesslike, systematic, *versus* evasive, unplanned, slipshod, teacher behaviour

Pattern Z — stimulating, imaginative, surgent, *versus* dull, routine, teacher behaviour (Ryans, 1960, 382).

Pattern Y and Pattern Z seemed to be analogous to two of the concepts discerned in the work of the Science Facilities Committee. Pattern Y seemed closely related to the notion of organization. In fact Rosenshine (1971) saw Ryans' Pattern Y as overlapping with both goal directed and organized behaviour.

Pattern Z was very similar to the classroom process variable which Anthony (1967) called Stimulation. Anthony's measure of stimulation included assessing variety in materials, techniques, and feedback to students. Keeves (1972) also used a variable called 'stimulation for learning' which incorporated measures of the variety of methods used.

Ryans' (1960) Pattern X did not quite correspond to the notion of involvement which was being sought. A much closer analogue was the second process variable which Anthony (1967) related to achievement: Self Involvement Opportunities. Keeves (1972) included two process variables similar to this: Affiliation with Class and Interaction with Students.

Hence, the three concepts discerned in the stated and implied intentions of the Commonwealth Standards Committee when designing rooms were also found in studies of teacher behaviour and classroom processes. The three major concepts around which the science room environment section of the questionnaire was structured were:

I Involvement: An analogue of Anthony's (1967) self involvement opportunities, as found in a classroom process study.

O Organization: An environmental analogue of Ryans' (1960) Pattern Y (businesslike behaviour) which was intended as a measure of the orderliness and convenience of the workplace in having space to work and being able to obtain the necessary apparatus and services when required.

V Stimulation through Variety: An environmental analogue of Ryans' (1960) Pattern Z but refined in the manner suggested by Anthony (1967) to mean variety in materials used and techniques of teaching.

Activities in Learning Science

The principal justification for building laboratories in schools, must surely be that students are thereby able to conduct experiments. It was therefore considered desirable to ascertain whether students enjoying better facilities than their peers in other schools would report different emphases in the science teaching which they experienced. There is little doubt that first hand experience and original investigations have become accepted as crucial elements in modern science curricula (Lee Dow, 1971; Ramsey, 1972).

Science rooms built under the aegis of the Australian Science Facilities Program were intended to accommodate and foster these teaching methods. Standards were set so that schools would have sufficient science rooms of suitable design to allow student experimental work, as appropriate, in any science lesson.

The development of instruments through which students can report upon the emphases in the science teaching was based upon the IEA *Description of Science Teaching* scales (Comber and Keeves, 1973). Modifications of them by Coxhead (1974) and by Kelly (1975) were also considered for possible use in this study.

The results of a trial questionnaire in Victorian schools prior to the main study suggested that the most promising approach was to use the items which had been used by the IEA. Two changes were made to the format. A five point, rather than a three point, response key was used. In addition to this, the instructions were modified so that a more specific meaning was given to each of the five categories.

The three factors which emerged from the analysis of the questionnaire were of direct relevance to the Science Facilities Project. They resembled the concepts of laboratory work, enquiry learning, and textbook learning around which the original IEA scales were formed but did not assume the same bipolarity. Scales which were able to be formed from these three factors had sufficiently good reliabilities to be used as measures of class activity in science. The interpretation of the factors and hence the attributes assumed to be measured by the scales were as follows:

E	Encouragement to Explore: the extent to which students are given encouragement to participate in active science learning in and outside school.
T	Textual: the degree to which science teaching is based upon textbooks.
P	Practical: the emphasis upon practical work in science classes.

Procedures

Collecting the information

The questionnaire, which was administered to students, contained the three environment scales and the three activities scales described. It was administered during the course of visits to 29 schools in addition to the interviews held with teachers and observations made of the use of facilities in teaching science.

Schools were chosen for this part of the Project so that the full range of standards of science facilities was presented. The selection was made from a larger sample of schools of which a survey of science facilities had been conducted (Ainley, 1976).

In developing the student questionnaire attention had been focussed upon Year 9 science teaching. It was inferred from other studies (Owen, 1975; Wilson, 1975) that facilities had a greater effect on junior science teaching than that in Years 11 and 12. In addition, the choice of one of the Junior Secondary Years was likely to achieve maximum variation in the facilities available to science classes.

In its final form, the student questionnaire was administered to almost 3000 Year 9 students in twenty-nine Australian schools. Within each school three or four Year 9 science classes were chosen at random so that a total of about 100 students from each school were included. Choosing classes seemed important because many of the attributes being studied were possibly more characteristic of classes than of schools. Such a choice made it possible to use the class as the unit of analysis.

Reliabilities were calculated for each scale (Involvement, Organization, Stimulation, Encouragement, Textual, and Practical) in two ways. Firstly, the values using the student responses were calculated. Secondly, the class means for each item were computed for each of the 104 classes and the reliability of each scale using class means was calculated. Reliabilities and other relevant statistics are presented in Table 1.

TABLE 1

Scale statistics for the student questionnaire
Views of science classrooms

	I	O	V	E	T	P
Number of items	7	7	7	9	6	6
Between Students (N = 2991)						
Reliability (Coefficient α)	0.59	0.75	0.55	0.64	0.66	0.66
Mean	196.	23.8	19.0	24.1	17.7	18.8
Standard Deviation	4.7	5.7	4.6	5.3	4.8	4.4
Between Classes (N = 104)						
Reliability (Coefficient α)	0.82	0.89	0.68	0.73	0.84	0.85
Mean	19.5	23.6	18.9	24.3	18.7	19.0
Standard Deviation	2.3	3.3	2.6	2.6	3.3	3.1

Note: I = Involvement
O = Organization
V = Stimulation through Variety
E = Encouragement to Explore
T = Textual
P = Practical

Measures of the availability and quality of science facilities

On the basis of a previously reported factor analysis (Ainely, 1976), it had been decided that school science facilities could be considered in four dimensions. These were the availability of rooms, the quality of those rooms, the abundance of apparatus, and the number of laboratory assistants. In studying the effect of facilities on the views students held of their science learning, four variables corresponding to these dimensions were used.

1. The occupancy of science rooms was measured simply as the percentage of science lessons held in science rooms.
2. The quality of the science room most often used was assessed by the school science co-ordinator. The measure of 'quality' was a score obtained by summing the school science co-ordinator's assessment of the availability and accessibility of 14 key features. The scores on that scale correlated well with an observer's rating (r = 0.83). The scale had been previously shown (Ainley; 1976) to discriminate between rooms considered to be of different quality and to have a satisfactory internal consistency (α - .68).
3. The abundance of apparatus in each school was represented as a composite score. Replies from science co-ordinators about the adequacy of major, minor and expendable apparatus did not correlate as expected with the criterion scores. Two possible explanations existed for this. Firstly, most schools regarded their apparatus as sufficient, and secondly, the responses were the opinion of the science co-ordinator. Consequently those responses were possibly related to the program which the school conducted. By using a procedure described by Tukey (undated) and Peaker (1975) a composite score, based on these raw data, was formed. A series of stepwise regression analyses of each scale against the apparatus variables was carried out. It appeared that

by taking the rating for minor apparatus less the rating for expendable apparatus a satisfactory predictor could be obtained. This procedure was tantamount to using the expendable apparatus rating as a means of controlling for the bias introduced by the co-ordinator who gave the ratings.

4. The provision of laboratory assistants was represented by the number of laboratory assistants per 1000 students in the school.

These four variables, which were used as measures of schools' science facilities, were considered to be potential predictors of each environment and activity score. They were entered in the appropriate regression analyses in a predetermined sequence. It was postulated that the first consideration ought to be the proportion of science lessons held in science rooms. As a second effect, the quality of those rooms was considered to be a potential influence on the students' views of their science environment. The third level to be considered was the abundance of apparatus and finally the provision of ancillary staff was entered as a predictor.

Results and discussion

Science learning environments

The results of the hierarchical stepwise regression analyses which were performed are shown in Table 2. It can be seen that for these measures of aspects of the learning environment – involvement, organization and stimulation – about 30 per cent of the variance could be predicted from variations in the facilities available. It can be noted that all four aspects of science facilities were associated with environments which gave students the opportunity to be involved in purposeful activity. Even though the number of laboratory assistants per 1000 students was not significantly correlated with involvement (when a prior correction was made for occupancy, room quality and the abundance of apparatus) it was associated with involvement. Science lessons were apparently better, in terms of these dimensions, when held in good science rooms with sufficient apparatus and with laboratory assistants available.

TABLE 2

The impact of science facilities on the science learning environment

				Scale					
Step Variable	Involvement			Organization			Stimulation		
	R	r	β	R	r	β	R	r	β
	.52			.59			.48		
1. Occupancy		.45	.26		.45	.10		.33	.11
2. Quality		.40	.22		.53	.45		.34	.18
3. Apparatus		.30	.22		.30	.22		.40	.31
4. Lab. Assistants		01	.16		.03	.19		-	-

Note: N = 104 classes

The relative size of the beta weights was used to infer the relative contribution made by each variable to the prediction of the environment scales. Good quality rooms, which were defined in terms of those in which services were readily accessible, seemed to be an important component of a well organized environment. By contrast, it was the abundance of apparatus which was associated most strongly with the degree of stimulation that arose through variety in materials and methods. All four components of facilities made comparable contributions to an environment through which students were involved in purposeful activity.

In brief, good facilities did seem to make a difference to the views which students had of their science environment. Greater involvement, better organization and more stimulating variety were reported by students who enjoyed better facilities.

Activities in science

To examine the combined impact of various aspects of facilities on activities, an hierarchical regression analysis was again performed. The same conceptual framework, as for the environment scales, was used to determine the sequence in which variables were to be entered in the regression analysis. The results for these analyses, which are shown in Table 3, were less conclusive than for the environment scales. However, there was an association between the facilities used and the students' reported activities. Practical work received greater emphasis in those classes which were held in science rooms and had abundant apparatus available in their school. By contrast, those classes whose science lessons were frequently in non-science rooms relied more on textual materials. In brief, the relative use of practical work and textual materials were associated with the proportion of lessons held in science rooms.

TABLE 3

The impact of science facilities on activities in learning science

Step Variable					Scale				
	Exploration			Textual			Practical		
	R	r	β	R	r	β	R	r	β
	.31			.26			.37		
1. Occupancy		.04	-.18		-.25	-.22		.30	.23
2. Quality		.17	.28		-.19	-.06		.21	.00
3. Apparatus		.16	.23		-.07	-.01		.30	.22
4. Lab. Assistants		.10	.23		-.02	-.06		-.12	.02

Note: N =104

For the scale which was designated as encouragement to explore the results were less clear. Only the quality of the room and the abundance of apparatus were correlated with the scale. However, the regression weights indicated that, while room quality, apparatus, and laboratory assistants were associated with more of this activity, the beta weight for occupancy was negative. An explanation for this would appear to be in the fact that many of the activities involve going out of the science room to such locations as the library. Moreover, it did seem to be an act of choice on the part of the science teacher or science staff to use the facilities in this way. Encouragement to explore was most strongly correlated with the number of science rooms per 1000 students (.18) and the average quality of the schools' science rooms (.24). Apparently it was a choice within a good suite of science rooms.

A glimpse at some schools

When visiting schools as part of the Science Facilities Project another task was performed. The schools were to form a set of case-studies of facility use. Hence science teachers and others on the staff were interviewed and, where possible, science classes were observed. The bulk of the case study data cannot be presented here, but two examples are particularly relevant to these results.

Petrin High School was a developing school. In 1975 it was visited at a time when science accommodation was one modestly equipped portable classroom and two other portables with no special equipment. An activity-based curriculum (ASEP and other material) was used but its use involved some difficulties for the staff. Storage provision was almost non-existent and services were most inadequate. The direct impact of these privations was that some units were not able to be offered, some activities in other units were omitted, and generally the range of choice was restricted. One teacher expressed the view strongly that this was particularly a problem in a school where students would benefit greatly from non-didactic teaching methods. The youthfulness and enthusiasm of the staff resulted in a perseverance with the course they wished to follow. When a return visit to the school was made one year later, staff were able to reflect on the difference the occupancy of a new suite of four modern, well-designed science rooms had made to their teaching of junior science. There was no dramatic change reported. However, teachers did report some more subtle differences. The most important was the convenience when working. Similar things could be done but more efficiently, and with less time wasted.

> I found a great time-saving though. It saved time because I've got things in the room. I don't have to carry them from room to room. I've got hot water on tap and ice here. . . Last year for hot water I had to send a thermos to the staff room to be filled from an urn. Well that takes 10 minutes or so. Now we just run a tap.

A further comment concerned the responsiveness of the students. It is quite possibly effects such as these which are reflected in the results of Table 2 and

> The kids are better in class because they feel they're in a science room and having everything you need is important . . . They seem more responsive. They look after equipment better. They're interested. You can get them into a pattern of knowing where to put things away and what to do. They know where things are, whereas they don't when you bring in five trays to the portable rooms. They forget what came off which tray and you have bedlam at the end.

Ruraton High School was a case notable for the full exploitation of its facilities. The four science rooms now in use were occupied in 1969. It was a valuable example because the present science co-ordinator was teaching at the school. He recalled the time prior to 1969 when science was in some very old rooms to which occasionally one might bring a bucket of water and a portable burner. Perhaps with the wisdom of hindsight he recalled that:

> . . . moving to the new building where there are four science rooms meant all classes could have all of their science in a science room, and in a particular room. Activity-based programs which we run now would not have been possible under the old situation.

The new methods were not entirely based on new curriculum materials from outside. Original material developed in the school was also used. The science co-ordinator has been at the school since 1960 except for a few years at other schools. He recalls that:

> Since the new rooms we started writing our own programs. Now some of these programs we think are very good indeed. And we may perhaps use them in preference to ASEP material because they've been written by people here, suit our needs and so on. But the kids have been doing ASEP type activities since the first full year we moved into the new building.
>
> Q: Are you suggesting that the new rooms were a stimulus to writing original material?
>
> A: Certainly. There was no point otherwise.

It was a view shared by the Acting Principal, himself a science teacher, who came to the school not long after the new block was occupied.

Two other comments are worth quoting. The first emphasizes that that school was not constrained by its new facilities to ignore other than experimental work.

> We try not to just run through a unit as is . . . we punctuate this with rather more formal demonstration lessons, discussion lessons if a unit is not available for an area of the course which we think is important or more importantly because we feel that every now and again the kids respond better to a change in treatment.

The second indicates the sort of enrichment which can be provided with good facilities.

> I don't necessarily think that to do worthwhile science you have to have a great amount of material but I think it nice to be able to draw on material if need be. Just to quote an example. Not being a physicist I don't know much about cathode ray oscilloscopes. But with the help of the Physics people, when I was doing some work on nerves we set up a good demonstration. If it didn't exist I wouldn't get upset. I'd survive but it's better to have it. And it's better to have a room with a few taps in it, a spacious room, a quiet room where desks can be moved into any shape that's required .

Interviews with teachers and general observations in these schools yielded information consistent with the results which were recorded in Tables 2 and 3. These schools were mentioned because they provide examples where it was possible to trace the effect of relatively recent improvements in science facilities. The comments suggest not so much a dramatic transformation of learning activities but of more subtle changes which are reflected in students' views. They suggest that the changes which occurred may be due to an interactive effect of:

1. removing barriers to activities which were always desired,
2. changing attitudes or motivation of teachers, and
3. changes in the responsiveness of students

The data which were used in this study do not allow these distinctions to be pursued.

Concluding remarks

It should be stated quite emphatically that these results in no way refute the proposition that 'good teaching can occur in any conditions'. There are sufficient examples of active involved learning occurring in poor science rooms to support that. What the results do suggest is that good facilities are one component which influences the views students hold of science lessons and the activities in which they engage. That result is consistent with Brown's (1977) comment that schools with poor science facilities were inclined to view the Scottish Integrated Science Scheme with disfavour. It is also in accord with the results reported by Englehardt (1968) in the United States, that good facilities promoted the use of enquiry methods. To that extent the provision of good science facilities has been of value. Good science facilities appear to have fostered the sort of developments in science teaching which have been held to be desirable in recent years.

Acknowledgment

Work on the preparation of this paper has been supported by a grant from The Schools Commission.

References

AINLEY, J.G. The Australian science facilities program: A study of its impact on schools. *Bulletin of the Victorian Institute of Educational Research*, 1976, 36, 17.

ANTHONY, B.C.M. The identification and measurement of classroom environmental process variables related to academic achievement. Unpublished Doctoral Dissertation, University of Chicago, 1967.

BROWN, S.A. A review of the meanings of, and arguments for, integrated science. *Studies in Science Education*, 1977, 4, 31-62.

COMBER, L.L. & KEEVES, J.P. *Science education in nineteen countries: International studies in evaluation I.* Stockholm: Almqvist and Wiksell, 1973.

COMMONWEALTH OF AUSTRALIA: Commonwealth advisory committee on standards for science facilities in independent secondary schools (Chairman, E.D. Gardiner). *The design of science rooms*. Canberra: Australian Government Publishing Service, 1973.

COMMONWEALTH OF AUSTRALIA: Commonwealth advisory committee on standards for science facilities in independent secondary schools (Chairman, E.D. Gardiner). *Final Report, 1976*. Canberra: Department of Education, 1976.

COMMONWEALTH OF AUSTRALIA. Commonwealth Parliamentary Debates:
House of Representatives.
Fairbairn, D.F. 5 May 1971 p.2665; Menzies, R.G. 19 May 1964 p.1985.

COXHEAD, P. An empirical study of some aspects of secondary school science involving 11 to 13 year olds. Unpublished Ph.D. Thesis, Cambridge University, 1974.

ENGLEHARDT, D.F. Aspects of spatial influence on science teaching methods. Unpublished Doctoral Dissertation, Harvard University, 1968.

KEEVES, J.P. *Educational environment and student achievement*. Stockholm: Almqvist and Wiksell, 1972.

KELLY, A. *Scales of school learning conditions,* IEA Data Bank Form 8-751117 (November), 1975.

LEE DOW, K.C. *Teaching science in Australian schools.* Melbourne: Melbourne University Press, 1971.

NOVAK, J.D. *Facilities for secondary school science teaching.* Washington: NSTA, 1973.

OWEN, J.M. *The effect of schools on achievement in science*, IEA (Australia) Report No 1975 : 1. Melbourne: Australian Council for Educational Research, 1975.

PEAKER, G.F. *An empirical study of education in twenty-one countries.* Stockholm: Almqvist and Wiksell, 1975.

RAMSEY, G.A. Curriculum development in secondary school science. *Quarterly Review of Australian Education*, 1972, 5(*1*) and 5(*2*).

ROSENSHINE, B. *Teaching behaviors and student achievement*. London: NFER, 1971.

RYANS, D.G. *Characteristics of teaching.* Washington: American Council on Education, 1962.

TUKEY, J.W. Rosettes as carriers of regression. An unpublished paper from Princeton University and Bell Telephone Laboratories. Undated.

WILSON, A.F. *The effects of schools in Victoria on the science achievement of junior secondary students*, IEA (Australia) Report 1975: 2, Melbourne: ACER, 1975.

Research in Science Education, 1977, 7, 223-230.

THE APPLICATION OF THEORETICAL RESEARCH FINDINGS TO A PRACTICAL LEARNING PROBLEM

Geoffrey W. Beeson

Introduction

An examination of the Proceedings of the ASERA Conference for the six years 1971-76 shows that there has been a strand of interest in investigating theories relevant to learning in science (e.g., White, 1972; West, 1973; Mackay, 1975). In these theoretical studies, the interest has been in establishing the validity of the theories. This is an important activity because (a) it can help in gaining an understanding of how learning occurs, and (b) it has the potential for improving the effectiveness of classroom teaching and learning through the development of appropriate procedures and materials which are consistent with the theories.

Although this type of research usually has involved the participation of school or college students, the learning involved has often occurred under rather special conditions, for example, learning individually or in small groups removed from the rest of the class, using unusually large amounts of testing, learning from specially prepared programs, covering work not part of the normal course, and having the work taught by a visiting researcher-teacher. Because of this, the extent to which the research findings can be generalized to "normal" classroom situations is severely limited. There is therefore a need for research designed to test more specifically the extent to which the theoretical findings can be applied to classroom learning in general, and in particular, to practical learning problems that occur in schools, colleges and universities. This would be a type of applied research (Carroll, 1968) in which theoretical findings from more basic research would be tested in particular situations such that extrapolation from the basic research would be risky.

The purpose of this paper is to outline a study designed to investigate the applicability of some theoretical research findings concerning intellectual skill learning to a practical learning problem. In doing this, attention will be focussed on certain problems which can arise in this type of applied research.

Research findings on which the study was based

Three aspects of theory and theoretical research formed the basis for the study. These were:

1. the identification of intellectual skills as a major category of learning outcome;
2. hierarchical learning of intellectual skills;
3. the rote-meaningful dimension in intellectual skill learning.

Categories of Learning Outcome

Gagné (1972) has argued that intellectual skills can be identified as one of five categories of learning outcome. The five categories are: motor skills, verbal information, intellectual skills, cognitive strategies and attitudes. An intellectual skills can be described as the ability to perform a class of tasks, such as adding integers, interpreteing graphs, using the principle of Conservation of Momentum, identifying electric circuits.

Hierarchical Learning of Intellectual Skills

Over the last 15 years there has been a considerable amount of research designed to investigate the concept of learning hierarchies advanced by Gagné (1962, 1968). He defined a learning hierarchy as a network of successively achievable intellectual skills which lead up to the performance of a more complex task. Comprehensive reviews of the research into learning hierarchies have been made by Gagné (1973) and White (1973). In general, the results of this research have provided strong support for the proposition that intellectual skills are learned hierarchically.

The Rote-Meaningful Dimension in Intellectual Skill Learning

A number of research studies have been prompted by the work of Ausubel (1968) concerning the role of the learner's prior knowledge in facilitating meaningful learning. A recent study by Beeson (1976) was concerned with investigating the learning of intellectual skills in relation to Ausubel's theory. The results of this study indicated that the knowledge context in which the skills were learned influenced how meaningfully they were learned. A recently completed study by West (in press) has provided support for these results.

Outline of the study

The study was concerned with learning problems which occur in the transition from secondary to post-secondary education. The post-secondary courses which were used as a basis for the study were the Technical and Further Education (TAFE) courses in Victoria designed to prepare middle level technicians for employment (Education Department of Victoria, 1977). These courses may be commenced after the completion of Year 11 of secondary education. Their normal duration is four years part-time or two years full-time and successful students are awarded a Certificate of Technology.

> The question to be answered by the study was essentially: "Can the theoretical research findings concerning the sequence and meaningfulness of intellectual skill learning be applied to the problem of overcoming specific deficiencies in the knowledge background of beginning TAFE middle-level students?"

Rationale for the study

Students come to Victorian TAFE middle level courses with a wide variety of background knowledge and experience. This is because:

(a) some students come direct from various secondary schools and others after some time in employment;

(b) curricula in secondary schools are no longer subject to detailed external prescription, and therefore there are wide variations from school to school;

(c) in conjunction with (b) above, there has been a reduction in the use of external examinations;

(d) there are always large individual differences among students.

This range of background knowledge presents problems for the TAFE middle level teacher in deciding at what level to start his/her course. No matter what level is chosen, some students will be disadvantaged. They will become bored when the starting level is too low and frustrated when the starting level is too high. In both cases, student motivation flags with the consequences that the general level of performance is not as high as it could be, absences increase and some students drop out.

Stages in the research

The following stages were planned for the study:

1. Select a sample of at least 100 students from five Melbourne Metropolitan TAFE colleges conducting middle level courses in Mechanical, Production and Civil Engineering. Select at least five teachers, at least one from each college, to take part in the research. (April, 1976)
2. In conjunction with the five teachers, select a topic or small group of topics from the first-year Physics course for Mechanical, Production and Civil Engineering which would be suitable for use as the basis for the study. (April)
3. Have the teachers of first-year Physics define the intellectual skills needed for students to begin the topics chosen. (May)
4. Develop instruments to test whether the beginning students possess these prerequisite skills. (June-July)
5. Test students for the possession of these skills. (end July)
6. Where there are deficiencies in preprequisite skills, develop learning hierarchies for the missing skills. (August)
7. Prepare individualized learning materials for the teaching of each of these prerequisite skills or groups of skills, basing the learning sequence on the hierarchies developed, and using an appropriate knowledge context, together with advance organizers as needed, to promote meaningful learning of the skills. Build in self-correcting tests so students can check their own progress. This step will result in the development of several short learning units. (August-September)
8. Make the appropriate learning units available to students who have deficiencies in prerequisite skills, so they can be used for individual study in the two weeks prior to the teaching of the nominated topic(s). (September-October)

9a. Request teachers to teach nominated topic(s) in the normal way, and *on the assumption that all students possess the prerequisite intellectual skills*. (October)

9b. Evaluate the effectiveness of the learning units on the basis of students' and teachers' responses to questionnaires. (September-October)

10. Evaluate the approach used. (November-December)

10b. Revise the learning units in the light of the evaluation of them. (Nov. 1976-Feb. 1977)

Stages 6, 7 and 10b were carried out by a working party of three teachers following an introduction to the relevant theory, and under the general guidance of the project director. The development of the learning materials by teachers was regarded as an essential part of the catch-up approach being investigated. Moreover, the involvement of practising teachers was seen to have the additional benefits of helping to make the research findings more credible in the eyes of teachers, and of helping in their dissemination. The study was known as MILSCUP (Middle Level Skills Catch-Up Project) and was supported by a grant from the Technical and Further Education Commission. Further details are described elsewhere (Beeson, 1977).

Some problems encountered

Authors of textbooks on educational research methods frequently give attention to general problems encountered in conducting research (e.g., Isaac and Michael, 1971). Campbell and Stanely (1963) have identified many possible sources of invalidity in their treatise on experimental design. Occasionally there are articles in research journals which are concerned with such problems. For example, Thomas (1974) has discussed some specific problems associated with survey research in a particular context. In a recent paper, Richey (1976) identified nine reasons for failure of experimental procedure and suggested some ways of avoiding such failures.

In the present study, some problems were encountered which were specific to the particular type of applied research which the study represented. The possible occurrence of most of these problems was predicted in the planning stages of the study. However, the extent of their influence on the research was not accurately predicted in many cases, even though a substantial amount of preliminary trialling had been carried out prior to the commencement of the study. These problems will now be identified and discussed briefly in the hope that this will sensitize researchers to their existence and lead to improved ways of overcoming them.

To a large extent the problems encountered were associated with the constraints imposed by the need to conduct the research within the normal operating framework of the institutions involved. They can be grouped under three main headings:

1. The research design
2. The role of the teachers involved in the study
3. Organization

The Research Design

Ideally, in order to attempt to answer the research question stated above, an experimental design involving a control group would be used. In such a design, students taking part in the study would be allocated at random to the experimental group, in which the catch-up program would be used, and a control group, in which the catch-up program would not be used. The performance of the students in the two groups would then be compared on the relevant criteria (achievement, interest level, attendance, drop-out rates and teaching and learning difficulties encountered). Such an experimental design is summarized in Figure 1 (Campbell and Stanley, 1963, p.195).

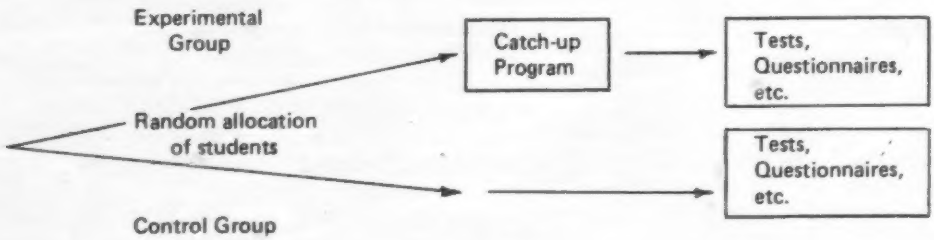

Experimental Group
Catch-up Program
Tests, Questionnaires, etc.
Random allocation of students
Tests, Questionnaires, etc.
Control Group

FIGURE 1

The preferred research design

Two major problems prevented the use of a control group design in the present study. In the first place, there were insufficient numbers of students available to make this design viable, especically in view of the high attrition rates known to exist in many TAFE courses.

Secondly, and more importantly, there existed the problem of allocating students at random to control and experimental groups. The particular problem was that the class, rather than the individual student, was the sampling unit, since by its nature the treatment (catch-up program) had to be carried out with whole classes. However, it was considered desirable to treat each individual as the unit of analysis. It is generally considered invalid to use individual scores from intact classes, since the individual scores are not independent events (Lord, 1967). One way of overcoming this difficulty would be to obtain a sufficiently large sample of classes to approximate the independent events requirement (Lord, 1963). It was planned that, when the number of classes available for the study was known, a statistical investigation be conducted to determine whether such an approximation could be made. However, as only nine classes were subsequently found to be available, this investigation was not carried out. Moreover, this number was insufficient for the class to be treated as the unit for analysis, and therefore the control-group design could not be applied.

The second choice for a research design was a quasi-experimental design: a non-equivalent control group design (Campbell and Stanley, 1963, p.217). However, this design could not be used because of the small and variable numbers of students in classes, the apparently great variability of student ability and motivation between classes, and the need to include a pre-test, which, in this study was part of the experimental treatment. Other quasi-experimental designs, such as a time series design (Campbell and Stanley, 1963, p.207) were also inapplicable. Because of these difficulties, a single group design, involving a trial of the experimental catch-up program followed by an evaluation, was developed. This design is summarized in Figure 2.

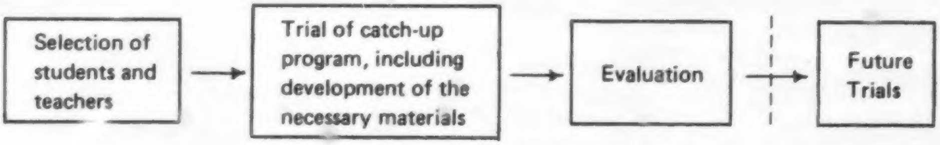

Selection of students and teachers
Trial of catch-up program, including development of the necessary materials
Evaluation
Future Trials

FIGURE 2

The research design used

The weakness of this design was that it did not allow a conclusive answer to be found to the research question from a single study, due to the inability to compare experimental and control groups. However, the design did allow evaluation of the catch-up program by students and teachers and it did provide for progressive improvements to be made to the program over a period of time. This latter would be achieved by using the results of the evaluation to modify parts of the catch-up approach where needed. This design is commonly used in Research and Development programs (Borg and Gall, 1971).

Two further problems were associated with the design of the research. These were the large attrition rates which were prevalent in the middle level courses, and difficulties associated with selecting teachers to take part in the study. Concerning the latter problem, the need to enlist the effective support of teachers and their colleges was essential. However, care had to be taken to ensure that teachers were selected whose overall level of co-operation with the project could somehow be regarded as representative of middle level physics teachers in general. In practice, selection was made following informal discussion with the teachers.

The role of the teachers involved in the study

The problems encountered under this heading included those associated with variations in the syllabuses in the different colleges, the need to ensure that the teachers involved followed the required "experimental" procedures, the importance of the teachers understanding the rationale of the study and the necessity for the teachers to learn some of the learning theory on which the study was based. The first three of these problems were aggravated by the fact that the study was conducted over a period of nine months.

Initially, it was necessary to select a group of topics from the first year Physics syllabus which would not be covered by any of the teachers involved before late September (see "Stages in the Research", above). There was some difficulty in making this selection due to variations in syllabus between colleges. Moreover, in some cases the teachers themselves had difficulty in keeping to their planned syllabus schedule, due to the pressures of time and other factors which affected their courses during the year.

It appeared to be self-evident that the extent to which a teacher would follow the required experimental procedure would be partly dependent on his understanding of the rationale for the study. Although at the commencement of the study all teachers had shown a sound grasp of the nature of the study and a general approval of its aims, at least two of the nine appeared to have lost this understanding later in the year when the "catch-up" materials were ready to be used. These teachers did not use the materials in a manner consistent with the aims of the project. It was concluded that a more extensive in-service education program for teachers would be needed in future trials.

There were two places where it was important for teachers to know some of the learning theory on which the study was based. Firstly, they had to be able to define prerequisite intellectual skills for the topics chosen (Stage 3 of the research). The results of the study indicated that the teachers could do this if they were provided with (a) a brief statement as to what sorts of learning are represented by the term "intellectual skills", together with several examples, and (b) a checklist of possible prerequisite skills.

Secondly, it was necessary for the three teachers responsible for writing learning materials (Stages 6 and 7) to have a working knowledge of the relevant theory sufficient to construct the materials in a manner consistent with the theory. To achieve this end, two two-hour learning sessions for which material was specially written, were conducted for the three teachers. In these sessions the emphasis was placed on (i) distinguishing intellectual skills from other types of learning outcomes, (ii) determining prerequisites for given intellectual skills, and (iii) developing appropriate contexts to encourage meaningful learning of the skills. No attempt was made to give the teachers practice in constructing complex, branching hierarchies, as this was considered to be unnecessary as well as unrealistic.

Organization

In this type of applied research study there can be considerable difficulties associated with the organization of the study due to the need to fit the research stages and procedures to the operating patterns of the institutions and individual teachers involved. In the present study these organizational difficulties were concerned mainly with supervision and timing.

Difficulties of supervision arose because of the decision that the experimental program in each college should be administered by the individual teachers involved rather than by the researcher. This decision was taken in order to avoid the introduction of special conditions into the learning situations. The variations in syllabus sequence already mentioned, and the differences in organization from one college to another, also made the monitoring of the project more difficult. Although difficulties of this sort had been predicted, they were of greater magnitude than expected. The existence of both day and evening classes, of part-time and full-time students, and of varying working times of the teachers involved all tended to complicate the day-to-day supervision of the study, making this a time-consuming aspect of the research.

Concerning timing, the most desirable time to trial the catch-up program was considered to be at the commencement of the first-year course (February), where deficiencies in background knowledge would be most noticeable. However, the timing of the study was affected by such external factors as the availability of funding and of students and teachers to take part in the study. The study had to be conducted over a period of time consistent with these outside influences, rather than at the time which was most desirable from a research design viewpoint.

Outcomes

The results of the study are discussed elsewhere (Beeson, 1977). Because of the nature of the research design used, one has to be cautious about drawing definite conclusions concerning the applicability of the research findings to the practical problem investigated. However, it can be stated with some certainty that:

(a) students did react favourably to the use of the individualized learning units to catch-up on prerequisite intellectual skills on which they were deficient;
(b) according to "armchair evaluators" and student reactions, teachers were able to write successful catch-up learning units based on the relevant theory;
(c) students considered it was reasonable to expect them to catch up on missing prerequisite skills in their own time;
(d) teachers from the four main colleges in which the learning units were used in accordance with the procedure required considered the approach helpful; all reported that it had made their teaching of the topics involved easier.

Conclusion

In this paper a study designed to test the applicability of some theoretical research findings to a practical learning problem has been outlined. Some problems encountered have been identified. It is suggested that more attention should be given to applied research of this type. Although the place of applied research in general has been well recognized in education, there has existed a gap between theoretical research findings and teaching and learning practices in educational institutions. Research of the type described could help to close that gap.

References

AUSUBEL, D.P. *Educational psychology: a cognitive view*. New York: Holt, Rinehart & Winston, 1968.

BEESON, G.W. Effects of context on hierarchical learning. Unpublished doctoral dissertation, Monash University, 1976.

BEESON, G.W. Overcoming specific deficiencies in the knowledge background of beginning TAFE students. Research Report, Technical and Further Education Commission, Canberra, 1977.

BORG, W.R. & GALL, M.D. *Educational research: an introduction*. (2nd ed.) New York: David McKay, 1971.

CAMPBELL, D.T. & STANLEY, J.C. Experimental and quasi-experimental designs for research on teaching. In N.L. Gage (ed.) *Handbook of research on teaching.* Chicago: Rand-McNally, 1963.

CARROLL, J.B. Basic and applied research in education: definitions, distinctions, and implications. *Harvard Educational Review*, 1968, **38**, 263-276.

EDUCATION DEPARTMENT OF VICTORIA. *Handbook of technical education 1977.* Melbourne: Technical Schools Publications Unit, 1977.

GAGNÉ, R.M. The acquisition of knowledge. *Psychological Review*, 1962, **69**, 355-365.

GAGNÉ, R.M. Learning hierarchies. *Educational Psychologist*, 1968, **6**, 1-19.

GAGNÉ, R.M. Domains of learning. *Interchange* (Toronto), 1972, **3**, 1-8.

GAGNÉ, R.M. Learning and instructional sequence. In F.N. Kerlinger (ed.) *Review of research in education* (Vol. 1), Itasca, Ill.: Peacock, 1973.

ISAAC, S. & MICHAEL, W.B. *Handbook in research and evaluation*. San Diego, Calif.: Knapp, 1971.

LORD, F.M. Elementary models for measuring change. In C.W. Harris (ed.) *Problems in measuring change.* Madison: University of Wisconsin Press, 1963.

LORD, F.M. A paradox in the interpretation of group comparisons. *Psychological Bulletin*, 1967, **68**, 304-305.

MACKAY, L.D. Cognitive preferences and achievement in physics, chemistry, science and mathematics. *Research in Science Education*, 1975, **5**, 49-58.

RICHEY, H.W. Avoidable failures of experimental procedure. *Journal of Experimental Education*, 1976, **45**, 10-13.

THOMAS, I.D. Survey research — some problems. *Research in Science Education*, 1974, **4**, 173-179.

WEST, L. Computerised Ausubel — support for Ausubel's theory from computer simulation studies. *Science education: research 1973*, 1973, **3**, 61-67.

WEST, L. Meaningful learning of intellectual skills. Research report, Education Research and Development Committee, Canberra (in press).

WHITE, R.T. Outline of a learning hierarchy investigation. *Research 1972*, 1972, 97-99.

WHITE, R.T. Research into learning hierarchies. *Review of Educational Research*, 1973, **43**, 361-375.

Research in Science Education, 1977, 7, 231-237.

CHANGES IN PUPILS' PERCEPTIONS OF SCIENTISTS, SCIENCE TEACHERS AND SELF

Lindsay D. Mackay and Richard T. White

Background

Among the outcomes that we hope for in teaching science in secondary schools is an increase in students' awareness of the attributes of scientists. We might expect that the longer students study science, the more accurate will be their knowledge of the desirable attributes for scientists to have in their work.

A second outcome we might hope for would be that students will come to accept the attributes of scientists as desirable models for themselves to adopt, and that ultimately they will see descriptions of scientists' attributes as applicable to themselves. We might expect that the longer students study science the more they will come to accept the attributes of scientists as their own.

We would expect science teachers to have a key role in developing students' perceptions of scientists, and in encouraging students to adopt the attributes of scientists as their own.

At last year's conference we reported the results of an exploratory study involving small numbers of students which was concerned with the effects of senior secondary school science courses and tertiary science courses on the attainment of the two objectives described above (White and Mackay, 1976). The results of that study led us to conclude that:

> "It seems that neither upper secondary and tertiary science courses nor cultural influences are doing anything to improve students' perceptions of desirable attributes of scientists.", and,
>
> "those who study science in forms 5 and 6 come to have self-perceptions less like scientists."

The present study

These results were sufficiently disquieting for us to commence a longitudinal study with a larger sample to further investigate the area. This paper is based on the results of the first testing of the longitudinal study. The first testing represents a cross-sectional survey of all students in Grades 9, 10 and 11 of three Victorian High Schools near the end of the 1976 school year.

Specifically, the analyses reported here attempt to answer five questions.

1. Do secondary school students in Grades 9, 10 and 11 differ in the accuracy of their knowledge of the desirable attributes of scientists?
2. Do students in Grades 9, 10 and 11 differ in the extent to which they see the attributes of scientists as applicable to themselves?
3. To what extent are students' self-perceptions congruent with their own perceptions of scientists and of science teachers?
4. Do students in Grades 9, 10 and 11 differ in the extent to which they perceive that science teachers possess the attributes that are desirable for scientists in their work?
5. To what extent are individual students' perceptions of science teachers congruent with their perceptions of scientists?

Knowledge of the attributes that are desirable for scientists, and perceptions of one's own possession of these attributes, may be measured by the Tests of Perception of Scientists and Self (Mackay and White, 1974, 1976). The two forms of the tests TOPOSS-Scientists and TOPOSS-Self were administered to the population of students in Grades 9, 10 and 11 in three randomly selected high schools in the Melbourne metropolitan area. These students were also administered a third test which measured students' perceptions of the extent to which science teachers possessed the attributes that are desirable for scientists – the Test of Perception of Science Teachers (TOPOST).

The tests were administered in one session in the order TOPOSS-Self, TOPOSS-Scientists, TOPOST. TOPOSS-Self was administered first to ensure that students' self-perceptions were obtained before they realised that the testing programmes had anything to do with science.

Results

Mean scores obtained on the three tests for students in the various schools and at the various grade levels are given in Tables 1, 2 and 3. These tables include summaries of analyses of variance. Table 4 provides the numbers in each of the groups in Tables 1, 2 and 3.

TABLE 1

ANOVA for scores on TOPOSS-Scientists

Source	df	MS	F
Grades	2	344.10	47.64**
Schools	2	200.54	27.77**
Grades x Schools	4	26.44	3.66*
Error	1198	7.22	

** $p < .001$
* $p < .01$

Mean scores on TOPOSS-Scientists

	School 1	School 2	School 3	Grade Total
Grade 9	15.65	13.99	16.10	15.25
Grade 10	15.99	15.86	17.10	16.32
Grade 11	16.90	16.77	17.56	17.14
School Total	16.18	15.54	16.99	

TABLE 2

ANOVA for scores on TOPOSS-Self

Source	df	MS	F
Grades	2	65.51	7.87**
Schools	2	56.17	6.75*
Grades x Schools	4	4.55	0.55
Error	1198	8.33	

** $p < .001$
* $p < .01$

Mean scores on TOPOSS-Self

	School 1	School 2	School 3	Grade Total
Grade 9	10.19	9.74	10.56	10.16
Grade 10	10.24	9.94	10.99	10.39
Grade 11	11.02	10.73	11.15	10.97
School Total	10.48	10.14	10.90	

TABLE 3

ANOVA for scores on TOPOST

Source	df	MS	F
Grades	2	160.86	17.35**
Schools	2	76.54	8.26*
Grades x Schools	4	15.80	1.70
Error	1198	9.27	

** $p < .001$
* $p < .01$

Mean scores on TOPOST

	School 1	School 2	School 3	Grade Total
Grade 9	13.81	12.42	13.54	13.26
Grade 10	13.76	13.04	14.11	13.64
Grade 11	14.53	14.40	14.63	14.52
School Total	14.03	13.29	14.09	

TABLE 4

Numbers of subjects in various groups in analyses in Tables 1, 2 and 3

	School 1	School 2	School 3
Grade 9	130	123	182
Grade 10	142	125	182
Grade 11	88	99	136

All three tests used consist of twenty pairs of statements, which are about scientists (TOPOSS-Scientists), about the respondent (TOPOSS-Self) or about science teachers (TOPOST). In each pair the statements have been shown to be equally socially acceptable to senior secondary school pupils, but to differ in their applicability to scientists. The respondent must choose the one of each pair which best describes the group concerned.

The TOPOSS-Scientists scores in Table 1 show a moderately accurate level of perception of desirable attributes of scientists. Since the respondent must choose one statement in each of twenty pairs of statements, someone with no perception at all of what scientists should be like can be expected to score about 10. In that light, a mean of between 15 and 17 out of 20 shows some general perception of what scientists should be like, but in our judgement not a particularly accurate one.

The ANOVA table and the mean scores for TOPOSS-Scientists in Table 1 indicate that there is a significant difference between students at the various grade levels in the accuracy of their perceptions of the attributes of scientists. Students in higher grades have more accurate perceptions of desirable attributes of scientists than students in lower grades. This trend is apparent for all three schools individually. Differences between the various schools in the accuracy of students' perceptions of desirable attributes of scientists, and the significant interaction between grades and schools suggests that schools may be having different effects in this area. However, conclusions on this must await the completion of the longitudinal study.

A mean of about 10 is to be expected for the TOPOSS-Self test when it is applied to a group that is not specially selected for its association with science, since the respondent has to choose one statement from each of twenty pairs. The ANOVA table in Table 2 indicates that students in higher grades have self-perceptions that are more like desirable attributes of scientists than do students in lower grades. However, for all groups, the TOPOSS-Self means are around 10 to 11, which indicates very little correspondence between students' self-perceptions and desirable attributes of scientists. Readers who believe that science courses should bring students to be more like scientists in their self perceptions will find little comfort in these results.

Mean scores on TOPOST in Table 3 lie about 13 to 14. For every group, students perceive that science teachers possess more of the desirable attributes for scientists than the students themselves possess.

The observation that students in higher grades perceived science teachers as possessing more of the desirable attributes of scientists than students in lower grades is heartening. However, the moderate mean scores suggest that science teachers are not perceived by students to possess the attributes of scientists to any great extent.

Congruence between perceptions of self, scientists and science teachers

Seventeen pairs of statements were common to the three tests TOPOSS-Self, TOPOSS-Scientists and TOPOST. Congruence in students' perceptions of the attributes of self, scientists and science teachers can be measured by the number of these pairs of statements in which the student selects the same statement as applicable to more than one group. This congruence measure takes no account of whether students' perceptions of the attributes of scientists are accurate or not.

Table 5 summarises the mean congruence scores on the 17 common statement pairs for students at various grade levels.

TABLE 5

Mean congruence scores for pairs of tests

Grade	Mean congruence score for:		
	TOPOSS Self / TOPOSS Scientists	TOPOSS Self / TOPOST	TOPOSS Scientists / TOPOST
9	9.77	11.14	11.55
10	9.73	11.26	11.87
11	10.16	11.73	12.67
Total	9.86	11.34	11.97

If there is no particular congruence between students' perceptions, a mean of about 8.5 would be expected in the congruence scores.

The results in Table 5 indicate that students' self-perceptions are not strongly congruent with their perceptions of the desirable attributes of scientists, as mean congruence scores for TOPOSS-Self and TOPOSS-Scientists lie around 10 out of a possible 17. There is no evidence of a consistent trend toward greater congruence in higher grades.

Congruence between students' self-perceptions and their perceptions of science teachers is measured by congruence scores for TOPOSS-Self and TOPOST. Mean congruence scores in this case lie around 11 to 12 out of a possible 17. Whilst students' self-perceptions are a little more congruent with their perceptions of the attributes of science teachers than with their perceptions of scientists, the congruence between self-perception and perception of science teachers is still not high.

Neither scientists nor science teachers appear to be perceived by students as possessing attributes which are desirable models for students to adopt, although students appear more ready to identify with the perceived attributes of science teachers than with those of scientists.

The greatest mean congruence scores in Table 5 for each grade level are those concerned with congruence between the perceived attributes of scientists and science teachers as measured by congruence between TOPOSS-Scientists and TOPOST. Mean congruence scores in this case are about 12 out of a possible 17. It appears that students do not perceive science teachers as particularly like scientists in terms of the desirable attributes of scientists included in the test. There is, however, a heartening trend for students in higher grades to perceive science teachers as more similar to scientists than do students in lower grades.

Correlations between scores on the tests

Correlations between scores on the three tests used in the study have been used to examine whether the accuracy of students' knowledge of desirable attributes of scientists (score on TOPOSS-Scientists) is correlated with

> the extent to which students perceive the attributes of scientists as applicable to themselves (score on TOPOSS-Self), and the extent to which students perceive that science teachers possess the attributes that are desirable for scientists (score on TOPOST).

The within schools correlations between these pairs of scores are provided in Table 6.

TABLE 6

Correlations between scores on the three tests

Grade	N	Within school correlations between:	
		TOPOSS-Scientists/TOPOST	TOPOSS-Scientists/TOPOSS-Self
9	435	.49	.20
10	449	.32	.17
11	323	.30	.13
Total	1207	.40	.20

There is a significant positive correlation over-all between scores on TOPOSS-Scientists and TOPOSS-Self, and between scores on TOPOSS-Scientists and TOPOST. It appears that accuracy of knowledge of the desirable attributes of scientists is positively correlated with the extent of self-adoption of these attributes, and with the extent to which science teachers are perceived to have the attributes of scientists.

There is, however, a significant trend for these correlations to be smaller for higher grade levels. It appears that accuracy of knowledge about scientists is more highly correlated with the extent of self-adoption of the attributes of scientists and with the extent to which science teachers are perceived as "scientific" in lower grades than in higher grades.

Conclusion

The results of this study suggest that the causes for disquiet identified in the exploratory study reported last year were well founded. The results of the longitudinal study of which this represents the first part should enable the issues to be explored in even greater depth.

References

MACKAY, L.D. & WHITE, R.T. Development of an alternative to Likert scaling. *Research in Science Education*, 1974, 4, 131-139.

MACKAY, L.D. & WHITE, R.T. *Tests of perception of scientists and self.* Melbourne: A.C.E.R., 1976.

WHITE, R.T. & MACKAY, L.D. The effect of science courses on perception of scientists and self. *Research in Science Education*, 1976, 6, 73-75.

THE CONTRIBUTORS.

JOHN AINLEY	Senior Research Officer, A.C.E.R.
GEOFFREY W. BEESON	Head, Dept. of Curriculum and Teaching, Rusden S.C.V.
EFFIE D. BEST	Senior Research Officer, S.A. Dept. of Education.
R. A. FAWNS	Lecturer, Faculty of Education, University of Melbourne.
BARRY J. FRASER	Lecturer in Education, Macquarie University.
PETER S. FREYBERG	Professor of Education, University of Waikato, N.Z.
P. L. GARDNER	Reader, Faculty of Education, Monash University.
J. K. GILBERT	Senior Lecturer, Institute of Educational Technology, University of Surrey, U.K.
JANET V. KEIGHTLEY	Research Officer, S.A. Dept. of Education.
N. C. KELLETT	Lecturer, HEARU, Monash University.
C. B. KINGS	Lecturer, Rusden S.C.V.
R. LEWIS	Lecturer, School of Education, Latrobe University.
P. P. LYNCH	Lecturer, Faculty of Education, University of Tasmania.
L. D. MACKAY	Assoc. Professor, Faculty of Education, Monash University.
M. N. MADDOCK	Senior Lecturer in Education, University of Newcastle.
SUE McBURNEY	Research Officer, CRIME, Dept. of Education, N.S.W.
B. C. NEWMAN	Senior Lecturer, School of Education, University of New South Wales.
JEFFREY R. NORTHFIELD	Lecturer, Faculty of Education, Monash University.
ROGER J. OSBORNE	Senior Lecturer, Physics Dept., University of Waikato, N.Z.
JOHN M. OWEN	Senior Research Officer, A.C.E.R.
RENATO SCHIBECI	Lecturer, School of Education, Murdoch University.

M. J. SULLIVAN	Research Officer, S.A. Dept. of Education.
DAVID J. SYMINGTON	Senior Lecturer, Toorak S.C.V.
FREDERICK J. THOMAS	Lecturer, School of Teacher Education, Riverina C.A.E.
L. H. T. WEST	Senior Lecturer, HEARU, Monash University.
RICHARD T. WHITE	Associate Professor, Faculty of Education, Monash University.
C. WOODRUFF	E.R.D.C. Research Fellow, Monash University.